Katedra Biochemii i Mikrobiologii, Instytut Biologii SGGW w Warszawie

AUTOREFERAT

Postępowanie habilitacyjne

Dr Małgorzata Dudkiewicz
2020-07-30

Spis treści

1. DANE OSOBOWE 3

2. POSIADANE DYPLOMY I STOPNIE NAUKOWE 3

3. INFORMACJE O DOTYCHCZASOWYM ZATRUDNIENIU, W TYM W JEDNOSTKACH NAUKOWYCH 3

4. OPIS OSIĄGNIĘĆ OKREŚLONYCH W ART. 219 UST. 1 PKT 2 USTAWY Z DNIA 20 LIPCA 2018 R. PRAWO O SZKOLNICTWIE WYŻSZYM I NAUCE 4

I. WSKAZANIE OSIĄGNIĘCIA WYNIKAJĄCEGO Z ART. 16 UST.2 USTAWY Z DNIA 14 MARCA 2003 R. O STOPNIACH NAUKOWYCH I TYTULE NAUKOWYM ORAZ O STOPNIACH I TYTULE W ZAKRESIE SZTUKI (DZ. U. NR 65, POZ. 595 ZE ZM.) 4

A). PODSTAWĄ UBIEGANIA SIĘ O TYTUŁ DOKTORA HABILITOWANEGO JEST ZBIÓR 6 PUBLIKACJI (SUMA IF = 25.57) OPATRZONY TYTUŁEM: 4

B) CEL NAUKOWY W/W PRAC I PODSUMOWANIE OSIĄGNIĘTYCH WYNIKÓW 6

1. Wprowadzenie

2. Opis metod analizy bioinformatycznej

2.1. Metody oparte na analizie sekwencji

2.2. Analiza strukturalna

3. Opis poszczególnych osiągnięć

3.1. Opracowanie metody obliczania różnic w energii wiązania kompleksów peptydów i receptorów komórek T do modeli różnych alleli cząsteczek HLA jako oceny znaczenia różnic allelicznych i antygenowych w MHC między biorcą i dawcą przeszczepu komórek hematopoetycznych.

3.2. Odkrycie domeny kinazowej z atypowym miejscem aktywnym w białkach z rodziny selenoprotein (SELO), obecnych m.in. u człowieka, sugerujące ich funkcję regulatorową w odpowiedzi na stres oksydacyjny.

3.3. Odkrycie nowej rodziny kinaz regulowanych jonami wapnia, Secretory Pathway Kinases (SPK), prawdopodobnie związanej z rozwojem schorzeń z grupy zaburzeń neurologicznych

3.4. Odkrycie bakteryjnych aneksyn prawdopodobnych przodków aneksyn eukariotycznych.

3.5. Przypisanie nieadnotowanej rodziny białkowej DUF2362 do klanu Macro – odkrycie potencjalnego nowego uczestnika komórkowego szlaku sygnałowego opartego na ADP-rybozylacji.

3.6. Potwierdzona laboratoryjnie identyfikacja nowego członka rodziny kinaz NFK3 – PEAK3/C19orf35: pseudokinazy inhibującej crkII. z funkcją regulatora procesu formowania się cytoszkieletu, związanej z procesami onkogenezy.

II. PODSUMOWANIE AKTYWNOŚCI NAUKOWEJ 34

1). Lista rozdziałów w monografiach naukowych i podręcznikach 34

[List of published chapters in scientific monographs] 34

2). Lista publikacji w czasopismach naukowych nieuwzględnionych w sekcji I.2., z podziałem na prace opublikowane przed i po uzyskaniu stopnia doktora. 34

3). Udział i prezentacje na konferencjach krajowych i międzynarodowych 38

4). Udział w komitetach organizacyjnych i naukowych krajowych lub międzynarodowych konferencji 39

5). Udział w realizacji grantów naukowo-badawczych 40

III. WDROŻENIA: 40

IV. PODSUMOWANIE ŁĄCZNEGO DOROBKU NAUKOWEGO 41

5. AKTYWNOŚĆ NAUKOWA REALIZOWANA W WIĘCEJ NIŻ JEDNEJ UCZELNI, INSTYTUCJI NAUKOWEJ, W SZCZEGÓLNOŚCI ZAGRANICZNEJ ORAZ STAŻE NAUKOWE 42

6. OSIĄGNIĘCIA DYDAKTYCZNE: 43

7. INNE INFORMACJE 44

1. Dane osobowe

Imię i nazwisko:

[Name]

Małgorzata Dudkiewicz

2. Posiadane dyplomy i stopnie naukowe

[Diplomas, degrees conferred in specific areas of science or arts, including the name of the institution which conferred the degree, year of degree conferment, title of the PhD dissertation]

a) doktor nauk biologicznych w zakresie biologii, specjalność: biologia, 2004,
Tytuł rozprawy doktorskiej: „Modelowanie presji mutacyjnej i selekcyjnej w genomie prokariotycznym", Uniwersytet Wrocławski; Wydział Nauk Przyrodniczych. Promotor: Prof. dr hab. Stanisław Cebrat.
b) magister ochrony środowiska, 2000, Zakład Genetyki Instytutu Mikrobiologii, Uniwersytet Wrocławski; Wydział Nauk Przyrodniczych.
Tytuł pracy magisterskiej: "Komputerowa symulacja ewolucji sekwencji kodujących", promotor: Prof. dr hab. Stanisław Cebrat.

3. Informacje o dotychczasowym zatrudnieniu, w tym w jednostkach naukowych

[Information on employment in research institutes or faculties/departments or school of arts]

a) 2004 - grudzień 2005: koordynator transplantacyjny w Dolnośląskie Centrum Transplantacji Komórkowych z Krajowym Bankiem Dawców Szpiku we Wrocławiu,
b) grudzień 2005 – 2008: adiunkt w Katedrze Biometrii Wydziału Rolnictwa i Biologii SGGW w Warszawie,
c) 2006 - 2014: konsultant w Centrum Organizacyjno–Koordynacyjnym ds. Transplantacji POLTRANSPLANT,
d) 2008 - 2019: adiunkt w Katedrze Doświadczalnictwa i Bioinformatyki, Wydział Rolnictwa i Biologii SGGW w Warszawie,
e) od 2014: kierownik zespołu Centralnego Rejestru Potencjalnych Niespokrewnionych Dawców Szpiku i Krwi Pępowinowej POLTRANSPLANT,
f) od 1 X 2019 r. adiunkt w Katedrze Biochemii i Mikrobiologii, Instytut Biologii SGGW w Warszawie.

4. Opis osiągnięć określonych w art. 219 ust. 1 pkt 2 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce

[*Description of the achievements, set out in art. 219 para 1 point 2 of the Act]*

-Lista osiągnięć naukowych lub artystycznych, które stanowią istotny wkład do rozwoju określonej dyscypliny *[List of scientific or artistic achievements which present a major contribution to the development of a specific discipline]*

I. Wskazanie osiągnięcia wynikającego z art. 16 ust.2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm.)
[INFORMATION ON SCIENTIFIC OR ARTISTIC ACHIEVEMENTS SET OUT IN ART. 219 PARA 1. POINT 2 OF THE ACT]

a). podstawą ubiegania się o tytuł doktora habilitowanego jest zbiór 6 publikacji (suma IF = 25.57) opatrzony tytułem:

Zastosowanie metod bioinformatycznych z uwzględnieniem modelowania struktur i kompleksów białkowych w analizie niescharakteryzowanych sekwencji o potencjalnym znaczeniu dla biologii medycznej.

AUTORZY	TYTUŁ	CZASOPISMO	IF*	PKT. WG MNiSW**
Dudkiewicz M., Malanowski P., Czerwiński J., Pawłowski K.	An Approach to Predicting Hematopoietic Stem Cell Transplantation Outcome Using HLA-Mismatch Information Mapped on Protein Structure Data.	Biology of Blood and Marrow Transplantation. 2009, (15), pp.1014-1025. DOI:10.1016/j.bbmt.2009.05.011	3,149	30
Indywidualny wkład w publikację: Koncepcja badań, projekt badań, wykonanie obliczeń, zebranie danych i analiza statystyczna, opis wyników, stworzenie publikacji.				
Dudkiewicz M., Szczepińska T., Grynberg M., Pawłowski K.	A Novel Protein Kinase-Like Domain in a Selenoprotein, Widespread in the Tree of Life.	PLOS ONE. 2012, 7(2), pp. e32138. DOI:10.1371/journal.pone.0032138	3,73	40
Indywidualny wkład w publikację: Wykonanie analiz i modelowania, interpretacja i opis wyników, tworzenie publikacji.				
Dudkiewicz M., Lenart A., Pawłowski K.	A Novel Predicted Calcium-Regulated Kinase Family Implicated in Neurological Disorders.	PLOS ONE. 2013, 8(6), pp. e66427. DOI:10.1371/journal.pone.0066427	3,534	40
Indywidualny wkład w publikację: Wykonanie analiz, modeli i rysunków, interpretacja i opis wyników, tworzenie publikacji				
Kodavali PK., Dudkiewicz M., Pikuła S., Pawłowski K.	Bioinformatics Analysis of Bacterial Annexins – Putative Ancestral Relatives of Eukaryotic Annexins.	PLOS ONE. 2014, 9(1), e85428. DOI:10.1371/journal.pone.0085428	3,234	40

Indywidualny wkład w publikację: Wykonanie analiz bioinformatycznych, udział w tworzeniu koncepcji pracy i tekstu artykułu, przygotowanie rysunków, tworzenie publikacji.				
Lopez M.L., Lo M., Kung J.E., Dudkiewicz M., Jang G.M., Von Dollen J., Johnson J.R., Krogan N.J., Pawłowski K., Jura N.	PEAK3/C19orf35 pseudokinase, a new NFK3 kinase family member, inhibits CrkII through dimerization	Proc Natl Acad Sci U S A. 2019;116(31): 15495-15504.	9,58	200
Indywidualny wkład w publikację: Wykonanie analiz bioinformatycznych, przygotowanie rysunków, współtworzenie publikacji.				
Dudkiewicz M, Pawłowski K.	A novel conserved family of Macro-like domains-putative new players in ADP -ribosylation signaling.	PeerJ. 2019;1(7):e6863. doi: 10.7717/peerj.6863.	2,35	100
Indywidualny wkład w publikację: Koncepcja badań, wykonanie analiz bioinformatycznych, modelowanie, stworzenie publikacji.				

** Wartość wskaźnika IF podawana jest wg zasady: rok publikacji zgodny z rokiem edycji Journal Citation Reports (JCR).*

***1)Publikacje od roku 2019 - WYKAZ czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych wraz z przypisaną liczbą punktów z dn. 31 lipca 2019: załącznik do komunikatu Ministra Nauki i Szkolnictwa Wyższego z dnia 31 lipca 2019 r.*

2) Publikacje od roku 2015 – Wykaz czasopism do Komunikatu MNiSW z 18.12.2015 w sprawie wykazu czasopism naukowych, z uwzględnieniem Komunikatu MNiSW z dnia 1 lipca 2016 r. o sprostowaniu komunikatu w sprawie wykazu czasopism naukowych wraz z liczbą punktów przyznawanych za publikacje w tych czasopismach

3) Publikacje od roku 2014 – Wykaz czasopism do Komunikatu MNiSW z 31.12.2014 w sprawie wykazu czasopism naukowych, z uwzględnieniem Obwieszczenia MNiSW z dnia 25 marca 2015 o sprostowaniu błędów

4) Publikacje od roku 2013 – Wykaz czasopism do Komunikatu MNiSW z 17.12.2013 w sprawie wykazu czasopism naukowych

5) Publikacje od roku 2011 – Wykaz czasopism do Komunikatu MNiSW z 21.12.2012 w sprawie wykazu czasopism naukowych

6) Publikacje z lat 2007-2010 – punktacja na podstawie Rozporządzenia Ministra Nauki i Szkolnictwa Wyższego z dnia 17 października 2007 r., w sprawie kryteriów i trybu przyznawania i rozliczania środków finansowych na działalność statutową (zgodnie z wykazem wybranych czasopism wraz z liczbą punktów za umieszczoną w nich publikację naukową z dnia 28 listopada 2007 r., z uzupełnieniami).

b) Cel naukowy w/w prac i podsumowanie osiągniętych wyników

1. Wprowadzenie

Dynamiczny rozwój nauk biologicznych w ostatnim okresie jest związany m.in. z zastosowaniem teoretycznej wiedzy o makrocząsteczkach oraz rozwojem metod opartych na informatyce, zarówno w analizach problemów ewolucyjnych, jak i w studiach poświęconych funkcjom poszczególnych białek tworzących proteom. Wykorzystanie nowych metod do opisu problemów biologicznych umożliwia analizę wielkich zasobów danych, przyrastających w tempie wykładniczym, generowanych codziennie przez tzw. laboratoria „high troughput" produkujące rocznie prawie 9 milionów nowych sekwencji DNA, szczególnie w przypadku analiz metagenomicznych, czyli sekwencjonowania tzw. próbek środowiskowych, zawierających DNA z czasem setek różnych organizmów. Nowe sekwencje stale powiększają zbiór znanych nauce białek, dodając wciąż nowe rodziny i klastry do już olbrzymich kolekcji. Różnica między liczbą białek badanych eksperymentalnie i liczbą sekwencji zgromadzonych w bazach danych ciągle rośnie zamiast się zmniejszać, co jest jednym z problemów współczesnej biologii. Sposobem na rozwiązanie tego problemu jest wykorzystanie metod opartych na podobieństwie sekwencji i struktur DNA oraz białek do wykrywania homologii między znanymi i nowopoznanymi sekwencjami w celu określenia funkcji, struktury, a nawet specyficzności substratowej produktów białkowych tych ostatnich. Mimo obserwowanego w ostatnich latach rozwoju metod i narzędzi informatycznych, udostępnianych często publicznie przez autorów, oraz zaawansowanych metod statystycznych liczba nie adnotowanych lub niewystarczająco opisanych sekwencji deponowanych w bazach danych rośnie. Jest to poważne wyzwanie, ale także szansa na nowe odkrycia i obserwacje. Wspólnym motywem przedstawionego poniżej cyklu publikacji jest m.in. wykorzystanie zaawansowanych metod badania homologii sekwencyjnej i strukturalnej do scharakteryzowania nieopisanych jeszcze rodzin białkowych.

Poszukując odległych homologii w przypadkach niescharakteryzowanych sekwencji lub rodzin sekwencji trzeba sięgać po metody wystarczająco czułe i specyficzne, aby wykryć odległe, ale prawdopodobnie spokrewnione adnotowane sekwencje w olbrzymich bazach danych, które pomogą sformułować hipotezę na temat funkcji i właściwości nowoodkrytych białek. Tradycyjne dopasowania sekwencji bazujące na klasycznych metodach przyrównania (alignmentu) pary lub wielu sekwencji (MSA-multialignment) oraz typowe heurystyki wyszukiwania lokalnych podobieństw sekwencyjnych w bazach danych (BLAST) są tu niewystarczające. Poszukiwanie odległych homologii bazuje na tzw. metodach profilowych, poszukujących podobieństw na poziomie porównania sekwencji i profili rodzin białkowych lub samych profili między sobą (sequence–profile, profile–profile i meta-profile comparison methods), z wykorzystaniem zaawansowanych modeli probabilistycznych typu HMM, (Hidden Markov Models), a także z wykorzystaniem metod przewidywania struktur drugorzędowych i innych cech wyższego rzędu, które można przypisać sekwencjom (Rozdział 2.1).

W następnym kroku przydatne są metody, które umożliwią zebranie argumentów i wyciągnięcie wniosków, np. modelowanie porównawcze (inaczej homologiczne) struktur, dokowanie potencjalnych substratów, analizowanie energii tworzenia kompleksów (Rozdział 2.2).

2. Opis metod analizy bioinformatycznej

2.1. Metody oparte na analizie sekwencji

Podstawowym narzędziem analizy bioinformatycznej jest algorytm dopasowania sekwencji, (sequence alignment). Dopasowanie polega na takim przyporządkowaniu symboli (nukleotydów lub aminokwasów) w dwóch porównywanych sekwencjach, aby otrzymać jak najwyższą wartość liczbową, która jest miarą podobieństwa sekwencji. Algorytm dopasowania globalnego polega na przyrównaniu dwóch sekwencji na całej ich długości i poszukiwaniu optymalnej ścieżki dopasowania przy zastosowaniu programowania dynamicznego, co w uproszczeniu oznacza dodawanie kolejnej pary znaków z przeliczeniem wartości dotychczasowego przyrównania przy każdym kroku i testowanie każdej możliwej opcji dopasowania znaków. Taki algorytm umożliwia oczywiście dopasowanie dowolnej pary sekwencji. Dlatego ważnym problemem badawczym jest ocena wiarygodności dopasowania, czyli stopnia, w jakim oddaje ono rzeczywiste pokrewieństwo ewolucyjne (homologię) porównywanych sekwencji. Miarą istotności dopasowania sekwencji jest tzw. expect value (e-value), której wielkość zależy od liczbowej wartości przypisanej danemu dopasowaniu. Ta z kolei zależy od przyjętej zasady punktowania trzech możliwych stanów dopasowania. Najpopularniejszymi w konwencjonalnej analizie bioinformatycznej systemami oceny wartości dopasowani sekwencji są macierze substytucji aminokwasowych: PAM (Percent Accepted Mutation) i BLOSUM (BLOcks SUbstitution Matrix), które powstały na podstawie substytucji obserwowanych w specjalnie wyselekcjonowanych zbiorach spokrewnionych sekwencji aminokwasowych kodujących białka pełniące te same funkcje u różnych organizmów. Obserwowane substytucje były z założenia akceptowane przez selekcję i nie wpływały na zmianę funkcji białek. Na podstawie tych obserwacji stworzono macierze 20 x 20 reprezentujące tendencje każdego aminokwasu do podstawienia przez inne. Wartości umieszczone w macierzy służą do oceny liczbowej stanów dopasowania sekwencji aminokwasowych, przy czym zarówno wartości zachowania poszczególnych aminokwasów (match), jak i różnych rodzajów substytucji (mismatch) są wyraźnie zróżnicowane. Od prostej macierzy uwzględniającej tylko substytucje można przejść do bardziej skomplikowanych modeli probabilistycznych wprowadzających tzw. „kary za przerwy" (gap penalties), przy czym z ewolucyjnego punktu widzenia uzasadnione jest zróżnicowanie kary za otwarcie i wydłużenie przerwy w dopasowaniu. W takim zaawansowanym modelu prawdopodobieństwa przejść między stanami zależą od stanu poprzedniego i od stanu bezpośrednich sąsiadów, co można opisać macierzą typu HMM (Hidden Markov Model). Takie macierze są zazwyczaj generowane dla rodzin białkowych bazy Pfam (Protein families). Algorytm dopasowania globalnego nie sprawdza się jednak w wypadkach, kiedy głównym celem analizy jest poszukiwanie sekwencji podobnych do kwerendy w ogromnych zbiorach danych, jakimi są bazy sekwencji biologicznych. Algorytm najpopularniejszego programu do przeszukiwania baz danych: BLAST (Basic Local Alignment Search Tool) opiera się na idei dopasowania lokalnego, które nie wymaga czasochłonnego poszukiwania optymalnej ścieżki dopasowania na całej długości przyrównywanych sekwencji, ale ogranicza się do poszukiwania krótkich wspólnych sekwencji tzw. „słów" (k-tup) w przeszukiwanym zbiorze sekwencji, a następnie wydłużaniu dopasowania w obu kierunkach aż do momentu, w którym całkowita wartość lokalnego dopasowania spadnie poniżej pewnej wartości progowej. Pozwala to na szybkie wybieranie z bazy danych sekwencji wykazujących lokalne podobieństwo do kwerendy i optymalizację dopasowań, a następnie ich ocenę statystyczną i prezentowanie list potencjalnych homologów uporządkowanych według kryteriów prawdopodobieństwa, że obserwowane dopasowanie nie jest

zjawiskiem czysto losowym. Ze zbioru potencjalnych homologów można stworzyć dopasowanie wielu sekwencji tzw. multialignment (MSA- Multi Sequence Alignment), który jest podstawą szeregu metod analizy bioinformatycznej wyższego stopnia: punktem wyjścia dla metod filogenetycznych, wykrywania zachowywanych wzorców i wyrażeń regularnych, przewidywania struktur drugo- i trzeciorzędowych białek lub identyfikacji reszt funkcyjnych.

W dopasowaniu wielu sekwencji odpowiadające sobie ewolucyjnie reszty są dopasowane w jednej kolumnie. W idealnej sytuacji reszty z tej samej kolumny MSA powinny zajmować te same pozycje w strukturze trójwymiarowej białka. Zwykle dzieje się tak jednak tylko w przypadku najbardziej trywialnych dopasowań. Przy poziomie identyczności sekwencji rzędu 30% tylko połowa reszt może być skutecznie dopasowana na poziomie struktury. Ponieważ sekwencja ewoluuje dużo szybciej niż struktura, nawet dostęp do danych strukturalnych nie gwarantuje bezwzględnej poprawności MSA. Najczęściej stosowanym algorytmem dopasowania wielu sekwencji są tzw. metody hierarchiczne (lub progresywne), wychodzące od macierzy odległości powstającej z porównania każdej możliwej pary sekwencji ze zbioru, a następnie stworzenia drzewa przewodniego na zasadzie analizy skupień, według którego kolejno dopasowuje się sekwencje, zaczynając od pary (par) o najmniejszej dywergencji, stopniowo dołączając coraz bardziej odległe. Algorytm progresywny bazuje na założeniu, że MSA może być traktowany jako zbiór dopasowań par sekwencji.

Kolejnym etapem rozwoju algorytmów budowania dopasowań wielokrotnych jest porównywanie sekwencji do profilu lub modelu probabilistycznego zamiast do surowej sekwencji, a w dalszej kolejności porównywanie profili i modeli między sobą, co pozwala na odkrycie bardzo odległych homologii, niedostrzegalnych na podstawowym poziomie analizy.

Metody wykrywania homologii bazujące na podobieństwie surowych sekwencji dwóch białek mają poważne ograniczenia, poziom identyczności między porównywanymi sekwencjami nie może spadać poniżej 25%. Wszystko co mieści się w okolicach tego progu jest zwyczajowo nazywane „twilight zone" bioinformatyki. Pokrewieństwo ewolucyjne wielu białek sięga właśnie tej „szarej strefy", ich detekcja wymaga więc metod o zwiększonej czułości. Nowe podejście do analizy sekwencji, które zwiększyło czułość wykrywania homologii wymagało zmiany definicji podobieństwa. Rodzina, czyli grupa białek pełniących te same lub zbliżone funkcje, może być scharakteryzowana za pomocą MSA zredukowanego do dwuwymiarowej tablicy, gdzie jeden wiersz odpowiada jednej kolumnie wielokrotnego dopasowania. Każda pozycja w kolumnie profilu opisuje prawdopodobieństwo wystąpienia danego aminokwasu w danej pozycji. Dodatkowe kolumny oznaczają pozycyjnie specyficzne kary za wprowadzenie przerwy, w ten sposób profil opisuje wszystkie dozwolone mutacje w każdej pozycji przyrównania stworzonego dla całej rodziny białkowej. Takich profili można użyć przy konstruowaniu algorytmu poszukiwania optymalnej ścieżki dopasowania sekwencji metodą programowania dynamicznego, stosując pozycyjnie specyficzny system wartościowania, ponieważ prawdopodobieństwa przejść (mutacji) zmieniają się w zależności od pozycji w sekwencji.

Naturalną konsekwencją opisanego powyżej podejścia do poszukiwania sekwencji homologicznych jest przejście na wyższy poziom, czyli porównywanie profili rodzin białkowych (modeli HMM) między sobą. Tak, jak w przypadku dopasowania dwóch sekwencji lub sekwencji do profilu, także tutaj ma zastosowanie metoda programowania dynamicznego i heurystyka optymalnego dopasowania lokalnego. Różnicą jest bardziej skomplikowany sposób wartościowania zestawienia profili, który musi się opierać na iloczynie wektorowym. Jeżeli

porównywanie profili ma służyć jako alternatywa do klasycznego wyszukiwania homologów, pojawia się jeszcze jeden poważny problem – zarówno kwerenda jak i baza powinny być przedstawione jako zbiór profili rodzin białkowych, co pociąga za sobą konieczność wykonania szeregu wstępnych dopasowań sekwencji i przeliczeń modeli. Popularnymi serwerami oferującymi dostęp do tego typu algorytmów są HHpred (Homology detection & structure prediction by HMM-HMM comparison)[1] oraz FFAS03[2] (Fold & Function Assignment Server), oba wykorzystywane w omawianych publikacjach własnych.

Porównanie dwóch modeli probabilistycznych znacząco podnosi czułość wykrywania homologów, ale wiąże się też z rosnącym zagrożeniem otrzymania fałszywie pozytywnych wyników, czyli niespokrewnionych sekwencji zidentyfikowanych błędnie jako homologi. Jest to zjawisko typowe tylko dla tej metody, ponieważ przy klasycznym porównaniu sekwencji otrzymujemy wynik prawidłowy lub brak wyniku. Stąd bardzo ważnym zagadnieniem w analizie odległych homologów jest znalezienie równowagi między dwoma efektami: wysoką czułością umożliwiającą uzyskanie nowych informacji, a narażeniem na błędnie pozytywne sygnały, które mogą prowadzić do błędnego rozpoznania homologii. Służą temu prawidłowa konstrukcja profilu, staranne przypisanie wag i dodanie informacji wynikającej z danych strukturalnych, o ile są dostępne. Takim udoskonaleniem opisanego powyżej podejścia są tzw. metody meta-profilowe (meta-profile comparison). Dodanie do profilu rodziny białkowej informacji strukturalnej (nawet w przypadkach, kiedy nieznana jest dokładna trzeciorzędowa struktura cząsteczki) poprawia znacząco czułość i specyficzność narzędzi bioinformatycznych. W praktyce oznacza to, że wartość przypisana przyrównaniu dwóch pozycji w macierzy programowania dynamicznego jest obliczana jako suma iloczynów wektorowych otrzymanych dla dwóch profili białkowych i dla dwóch modeli probabilistycznych opisujących strukturę drugorzędową porównywanych rodzin. Informacja o strukturze drugorzędowej może być też dodana ręcznie do dopasowania, np. na bazie konsensusu otrzymanego z różnych algorytmów przewidywania struktur drugorzędowych białka.

Jeszcze innym rodzajem problemu dla badaczy odległych homologii jest rozmiar rodzin białkowych, z którymi trzeba porównać sekwencje będące przedmiotem zainteresowania. W przypadkach, kiedy liczba potencjalnych relacji i podobieństw przekracza tysiąc, tworzenie wielokrotnych dopasowań i drzew filogenetycznych zaczyna być obliczeniowo zbyt wymagające, błędy kumulują się, a uzyskany obraz traci rozdzielczość. W celu analizy relacji między dużymi zbiorami sekwencji został opracowany algorytm CLANS (Cluster Analysis of Sequences)[3] dostępny w postaci aplikacji Java. Algorytm CLANS używa wariantu grafu Fruchtermana i Reingolda[4] do wygenerowania graficznej reprezentacji podobieństw między każdą możliwą parą sekwencji w dużym zbiorze. Sekwencje stanowią wierzchołki grafu, podczas gdy dopasowania lokalne są pokazane jako łączące je krawędzie o określonej sile przyciągania proporcjonalnej do wartości dopasowania. Aby zapobiec „zapadnięciu się" grafu, czyli skurczeniu się

---

[1] Meier A, Söding J. Automatic Prediction of Protein 3D Structures by Probabilistic Multi-template Homology Modeling. PLoS Comput Biol. 2015 Oct 23;11(10):e1004343. PMID: 26496371

[2] Jaroszewski L, Li Z, Cai XH, Weber C, Godzik A. FFAS server: novel features and applications. Nucleic Acids Res. 2011 Jul;39(Web Server issue):W38-44. doi:10.1093/nar/gkr441. PMID: 21715387

[3] Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics. 2004 Dec 12;20(18):3702-4. PMID: 15284097.

[4] Fruchterman, TMJ, Reingold EM. Graph Drawing by Force-Directed Placement. Software – Practice & Experience, Wiley, 1991, 21 (11): 1129–1164, doi:10.1002/spe.4380211102

do jednego wierzchołka między sekwencjami wprowadzono także słabe siły odpychania. Po losowym rozmieszczeniu sekwencji w przestrzeni dwu- lub trójwymiarowej następuje ewolucja grafu – wierzchołki tworzą skupienia i grupy, aż do momentu, kiedy siły przyciągania zostaną zneutralizowane przez siły odpychające, a graf zaczyna oscylować bez istotnej zmiany położenia jego punktów. Taka analiza jest często punktem wyjścia lub wskazówką do zauważenia związków między odległymi rodzinami białek.

2.2. Analiza strukturalna

Białka, których pokrewieństwo ewolucyjne jest bliskie, wykazują podobieństwo nie tylko na poziomie sekwencji, ale także na poziomie strukturalnym. Sekwencja ewoluuje dużo szybciej niż struktura, więc można założyć, że jeżeli obserwujemy podobieństwo na poziomie sekwencji, to także na poziomie struktury trzeciorzędowej dwa białka będą do siebie podobne. W konsekwencji można budować model struktury nieznanego białka w oparciu o znaną strukturę jego homologa. W najprostszych przypadkach, kiedy sekwencje różnią się w niewielu pozycjach, model jest prostą wizualizacją informacji, którą zawiera już dopasowanie między sekwencją interesującego nas białka, a sekwencją jego homologa o znanej strukturze, czyli tzw. matrycy. W przypadku odległych homologii model jest źródłem dodatkowych cennych informacji, których nie można uzyskać mając do dyspozycji tylko MSA, jak np. analiza interakcji pomiędzy odległymi na poziomie sekwencji, a bliskimi przestrzennie elementami struktury, obserwacja migracji centrum aktywnego, określenie zakresu możliwej interakcji między zmiennymi pętlami itp. Modele trójwymiarowe mogą także pomóc w przewidywaniu tworzenia kompleksów między różnymi białkami. Nawet same próby modelowania odległych homologów mogą stanowić użyteczne narzędzie weryfikacji hipotezy. Modele, które nie przechodzą walidacji, mają poważne problemy strukturalne lub brakuje im regionów ważnych dla stabilności cząsteczki, sugerują błędne dopasowanie i fałszywą hipotezę o homologii między interesującym nas białkiem, a sekwencją wybraną jako matryca do modelowania. W przypadkach, kiedy nie dysponujemy homologiczną matrycą na bazie eksperymentalnie rozwiązanej struktury, jedynym rozwiązaniem jest modelowanie z wykorzystaniem tzw. metod *„ab initio/de novo"*. Zakładają one, że można rozwiązać strukturę białka, znając jedynie jego sekwencję aminokwasową przez symulację fizycznych procesów prowadzących do zwijania natywnej cząsteczki. Niestety z powodu wciąż niedoskonale opisanych funkcji energetycznych i ogromnej liczby możliwych stanów konformacyjnych do przetestowania, tylko niewielkie białka (<120 reszt aminokwasowych) mogą być wiarygodnie modelowane w ten sposób.

Siły utrzymujące na swoich pozycjach atomy prawidłowo zwiniętej, biologicznie aktywnej makrocząsteczki wynikają z oddziaływań van der Waalsa, wiązań wodorowych, oddziaływań elektrostatycznych, oddziaływań hydrofobowych i ewentualnych wiązań bisulfidowych między resztami cysteinowymi. Dzięki rozbiciu tych sił na składowe jest możliwe obliczenie energii stabilizacji struktury makrocząsteczki (np. białka), jako różnicy poziomu energii swobodnej między formą zwiniętą a nieustrukturyzowaną tej samej cząsteczki (łańcucha polipeptydowego). Tworząc model pojedynczej cząsteczki lub kompleksu, należy wziąć pod uwagę wszystkie składowe i przeliczyć wartość energii stabilizacji dla danej konformacji. Większość pakietów programów służących do modelowania i analiz strukturalnych tworzy listę rankingową struktur/konformacji; od tych najkorzystniejszych energetycznie, do coraz niżej punktowanych.

Zadaniem bioinformatyki jest stworzenie jak najbardziej wiarygodnego modelu, który oddawałby lub przybliżałby z dużą dokładnością natywną strukturę białka. W ciągu ostatnich dziesięcioleci modelowanie białek przeszło istotną ewolucję od prostego dopasowania sekwencji docelowej do matrycy, do zaawansowanych metod hybrydowych. Wszystkie obecnie stosowane metody modelowania porównawczego (homologicznego) opierają się na kilku zasadniczych krokach:

- wybór matrycy do modelowania;
- dopasowanie sekwencji modelowanej do sekwencji matrycy;
- modelowanie zachowanego rdzenia białka;

- modelowanie regionów zmiennych (pętli);
- optymalizacja łańcuchów bocznych;
- rafinacja (dopracowanie modelu);
- walidacja modelu.

Dwa pierwsze punkty są kluczowe dla dokładności otrzymanego modelu, ponieważ błędów popełnionych przy wyborze matrycy i dopasowaniu sekwencji nie da się usunąć na kolejnych etapach modelowania. Są one również ściśle ze sobą związane, ponieważ poszukiwanie matrycy polega na przyrównaniu sekwencji, której strukturę chcemy modelować do wszystkich dostępnych sekwencji o znanej strukturze w bazach danych.

Metody poszukiwania matrycy możemy podzielić na czysto sekwencyjne (opisane powyżej metody profilowe i meta-profilowe lub alternatywne, czyli tzw. threading (przewlekanie) lub inaczej odwrócone zwijanie (inverse folding). W tej ostatniej metodzie dana sekwencja aminokwasów jest przeciągana przez znaną strukturę odtwarzając jej kształt, a wynik jest oceniany za pomocą odpowiedniej funkcji wartościującej lub wartości energii potencjalnej cząsteczki, opierając się na informacji uzyskanej na bazie znanych struktur rozwiązanych doświadczalnie. Z połączenia tych dwóch podejść powstały metody hybrydowe, które wykorzystują zarówno struktury potencjalnych matryc jak i wartości pozycyjnie specyficznych dopasowań sekwencji. Pozycyjnie specyficzna macierz substytucji używana do obliczania wartości dopasowania powstaje tu przez zastąpienie łańcuchów bocznych każdej reszty w strukturze przez łańcuchy wszystkich możliwych aminokwasów i obliczenie wartości dla poszczególnych substytucji przy użyciu statystycznie wyliczonych energii oddziaływań między resztami. Na jeszcze wyższym poziomie zaawansowania są tzw. metody meta-search, w których wykorzystanie podejścia statystycznego istotnie poprawia jakość dopasowania matrycy. Używa się tu wielu metod jednocześnie, a następnie analizuje się wyniki pod względem zgodności i możliwości wygenerowania konsensusu.

Wynikiem wyszukiwania matrycy może być więcej niż jedna struktura, przy uwzględnieniu ich nałożenia, możliwe jest stworzenie modelu na bazie dwóch lub więcej różnych struktur matrycowych reprezentujących osobne domeny.

Przed przystąpieniem do budowy modelu sekwencje modelowanego białka i matrycy powinny być do siebie optymalnie dopasowane. Takie dopasowanie nie powinno opierać się tylko na dwóch sekwencjach. Najczęściej do zbioru najlepszych matryc, zwróconych przez meta-serwer, przyrównywany jest MSA dla rodziny białkowej sekwencji, która jest przedmiotem modelowania. Procedura dopasowania wymaga często wprowadzenia manualnych poprawek i dokładnego sprawdzenia, czy algorytm nie wprowadził przerw, które są nie do zaakceptowania z biologicznego punktu widzenia oraz czy zgadzają się (przynajmniej w przybliżeniu) przewidywania dotyczące struktur drugorzędowych modelowanej sekwencji i struktury drugorzędowe przypisane do dopasowanych reszt w strukturze matrycy.

Dysponując optymalnym dopasowaniem pomiędzy MSA dla rodziny modelowanego białka i dla potencjalnych matryc, można określić regiony zachowane (conserved regions), w których występują te same reszty, zarówno w MSA, jak i w dopasowaniu struktur. Regiony te są uznawane za rdzeń modelowanego białka, którego konformacja pokrywa się z matrycą. Pozostałe fragmenty, w których obserwuje się dużą zmienność w dopasowaniu, lub których dopasowanie nie obejmuje, nazywane regionami zmiennymi, należą zwykle do pętli powierzchniowych lub dodatkowych elementów struktur drugorzędowych, które determinują specyficzność substratową lub pełnią funkcje regulatorowe. Cechy struktur matrycowych, takie jak sąsiedztwo i ekspozycja na kontakt z rozpuszczalnikiem, kąty i odległości między atomami w strukturze, oddziaływania Van der Waalsa i wiązania wodorowe służą do kalkulacji funkcji gęstości prawdopodobieństwa dla prostych warunków (więzów), jakie musi spełniać modelowana struktura.

Modelowanie pętli, czyli zmiennych regionów bez określonej regularnej struktury drugorzędowej, polega albo na modelowaniu *ab initio*, czyli zwijaniu krótkich peptydów, albo opiera się na tzw. bibliotekach, czyli zbiorach konformacji obserwowanych w naturze dla danego oligopeptydu. W przypadku regionów zmiennych o ustalonej strukturze, dla których nie ma odpowiednika w strukturze matrycy można zastosować modelowanie *ab initio* z więzami dla krótkich peptydów. Modelowanie dłuższych fragmentów wymaga bardziej zaawansowanego podejścia. Koordynaty reszt modelowanego białka w przestrzeni trójwymiarowej są generowane na bazie koordynat dopasowanych reszt matrycy, w przypadku łańcucha C-alfa taka zgodność nie powoduje zbyt wielu problemów, ale dla łańcuchów bocznych modelu może oznaczać liczne konflikty przestrzenne i konformacje odległe od optymalnych. Konformacje łańcuchów bocznych aminokwasów są określane przez kąty torsyjne (najwyżej 4, zależnie od długości łańcucha bocznego), zwykle przyjmują one tylko 4-6 możliwości konformacyjne, zwane rotamerami. Optymalizacja polega zwykle na przeliczeniu wariantów energetycznych dla kolejnych rotamerów danej reszty, co dla długich cząsteczek jest dość skomplikowane obliczeniowo.

Nawet po przejściu wszystkich wymienionych wyżej etapów, otrzymany model może być wciąż odległy od natywnej struktury, której wewnętrzne reszty są zwykle ciasno upakowane i osiągają lokalne minimum energii swobodnej, podczas gdy model może zawierać konflikty przestrzenne i puste kieszenie wewnątrz struktury, a jego C- i N-końce zwykle przyjmują losową konformację, bez zakotwiczenia w strukturze matrycy. Do rozwiązania powyższych problemów służą metody udoskonalenia (rafinacji) modelu (model refinement), które mogą bazować na zgromadzonej wiedzy o innych strukturach lub bezpośrednio na zasadach fizycznych. Metody oparte na prawach fizyki biorą pod uwagę podstawowe interakcje między atomami i próbują zmienić ich upakowanie w strukturze, aby spełniały określone warunki. Metody oparte na wiedzy wykorzystują potencjały statystyczne i informacje o matrycy zawarte w bazie PDB (Protein DataBank). W celu uzyskania lepszego modelu można strukturę poddać symulacji dynamiki molekularnej, co w ograniczonym zakresie może pomóc ustabilizować model i znaleźć najbliższe lokalne minimum energetyczne.

Ostatnim etapem opracowania modelu struktury białka jest jego ewaluacja. Metody tzw. „model assesment„ można podzielić na dwie grupy: metody statystyczne i bazujące na funkcjach energetycznych. Metody statystyczne bazują na grupowaniu podobnych modeli w klastry, a ich wyniki zależą od statystyki populacji. Podejście statystyczne zakłada, że największy i zawierający najbardziej kompletne modele (w sensie ilości modelowanych reszt) klaster ma najwyższą jakość.

Metody oparte na funkcjach energetycznych używają formuł fizycznych do oceny modelu i odróżnienia nierealistycznego modelu od zbliżonego do natywnego lub wykrycia ewidentnych błędów w modelowaniu struktury. Oprócz podania ogólnej oceny modelu, metody te umożliwiają także indywidualną ocenę poszczególnych reszt, co pomaga w wykrywaniu tzw. „hot spots" i usuwanie ewentualnych błędów lub nawet łączenie najlepszych fragmentów z różnych modeli w jedną strukturę.

3. Opis poszczególnych osiągnięć

3.1. Opracowanie metody obliczania różnic w energii wiązania kompleksów peptydów i receptorów komórek T do modeli różnych alleli cząsteczek HLA jako oceny znaczenia różnic allelicznych i antygenowych w MHC między biorcą i dawcą przeszczepu komórek hematopoetycznych.

An Approach to Predicting Hematopoietic Stem Cell Transplantation Outcome Using HLA-Mismatch Information Mapped on Protein Structure Data[5].

Wynik przeszczepienia komórek krwiotwórczych zależy w dużym stopniu od poziomu doboru dawcy i biorcy pod względem genetycznej zgodności antygenów HLA (Human Leukocyte Antigen). Procedura doboru nie zawsze kończy się wytypowaniem dawcy w pełni zgodnego, czyli zgodnego we wszystkich antygenach transplantacyjnych z potencjalnym biorcą (według obowiązującego obecnie m.in. w Polsce standardu jest to 5 par antygenów kodowanych przez loci HLA-A,B,C, DRB i DQB1). W przypadkach, kiedy nie można znaleźć w pełni zgodnego dawcy (tzw. dawcy 10/10), w praktyce klinicznej akceptuje się do przeszczepienia dawców z pojedynczą niezgodnością, przy czym nie jest do końca poznane, która niezgodność jest mniej szkodliwa dla profilu przeszczepienia, a tym samym akceptowalna dla transplantologa.

Celem pracy było modelowanie i ocena interakcji pomiędzy białkami kodowanymi przez różne allele HLA u biorcy przeszczepu, a receptorami limfocytów T i komórek NK (Natural Killer) układu immunologicznego dawcy. Taki typ interakcji prowadzi często do manifestacji objawów choroby przeszczep przeciwko gospodarzowi (GvHD: Graft versus Host Disease). Aktywowane cytotoksyczne limfocyty T (CTL- Cytotoxic T Lymphocytes) dawcy i komórki NK produkują granzymy, które wywołują reakcję GvHD. Celem pracy była analiza interakcji powodującej aktywację komórek CTL i NK na poziomie molekularnym, ze szczególnym uwzględnieniem roli HLA w działaniu tzw. synapsy komórkowej. Polimorfizm sekwencji alleli HLA klasy I i II decyduje też o pojawieniu się efektów host-versus-graft (gospodarz przeciwko przeszczepowi) i graft-versus-leukemia (przeszczep przeciwko leukemii). Wszystkie te efekty decydują o ostatecznym rezultacie przeszczepienia. Długoletnie obserwacje statystycznego przeżycia biorców przeszczepu komórek krwiotwórczych sugerują, że optymalne wyniki uzyskuje się przy maksymalnym dopasowaniu dawcy i biorcy, czyli na poziomie 10/10, ale wskazują także na różnice w przeżyciu w przypadku niezgodności dotyczących rożnych loci lub nawet antygenów/alleli występujących w obrębie tego samego locus. Analizy asocjacji statystycznych pomiędzy przeżyciem biorcy HSCT a rodzajem niezgodności HLA między dawcą a biorcą wykazują, że

[5] Dudkiewicz M, Malanowski P, Czerwiński J, Pawłowski K. Biol Blood Marrow Transplant. 2009 Sep;15(9):1014-25. doi: 10.1016/j.bbmt.2009.05.011,PMID: 19660714.

różnice zarówno na poziomie serologicznym (antygenowym), jak i allelicznym w zakresie wszystkich antygenów HLA klasy I oraz DRB1 w klasie II mają istotny wpływ na wyniki przeszczepienia[6,7]. Także podwójne niezgodności, niezależnie od loci były skorelowane z podwyższonym ryzykiem dla biorcy. Na bazie tych spostrzeżeń stworzono zbiory wytycznych funkcjonujących w praktyce klinicznej[8], które mają pomagać w podjęciu decyzji o wyborze dawcy do przeszczepienia. Do najważniejszych zasad należą:

-unikanie podwójnych niezgodności

-preferowanie niezgodności na poziomie raczej allelicznym niż antygenowym

-wybór niezgodności związanej z mniejszą liczbą podstawień aminokwasowych między allelami

-unikanie niezgodności, które są związane z ryzykiem rozwoju GvHD (jeżeli u biorcy występuje antygen nieobecny u dawcy)

-preferowanie niezgodności HLA związanych z substytucją poza rowkiem wiążącym prezentowany antygen.

Ta ostatnia zasada wynika z obserwacji opublikowanych już w 2001[9], które sugerowały, że substytucje w bardzo specyficznych pozycjach sekwencji aminokwasowej HLA mogą znacząco wpływać na wyniki transplantacji szpiku od niespokrewnionego dawcy. Szczególną rolę przypisano pozycji 116 w dłuższym łańcuchu HLA klasy I. Była to pierwsza próba analizy niezgodności HLA z uwzględnieniem danych strukturalnych. Od roku 2001 liczba eksperymentalnie rozwiązanych struktur zarówno dla samych antygenów HLA, jak i dla całych kompleksów TCR-pMHC (T-cell receptor- peptide- Major Histocompatibility Complex) znacząco wzrosła. Pozwoliło to na wybór trzech struktur kompleksów z bazy PDB, które zawierały zarówno HLA jak i prezentowany peptyd wraz z receptorem limfocytu T. Zostały one wybrane ze względu na interesującą właściwość, dotyczyły pary alleli B*35:01 i B*35:08, które w kontekście tego samego peptydu i TCR różniły się znacząco jeżeli chodzi o wywoływanie reakcji

---

[6] Flomenberg N, Baxter-Lowe LA, Confer D, Fernandez-Vina M, Filipovich A, Horowitz M, Hurley C, Kollman C, Anasetti C, Noreen H, Begovich A, Hildebrand W, Petersdorf E, Schmeckpeper B, Setterholm M, Trachtenberg E, Williams T, Yunis E, Weisdorf D. Impact of HLA class I and class II high-resolution matching on outcomes of unrelated donor bone marrow transplantation: HLA-C mismatching is associated with a strong adverse effect on transplantation outcome. Blood. 2004 Oct 1;104(7):1923-30. PMID: 15191952.

[7] Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, Eapen M, Fernandez-Vina M, Flomenberg N, Horowitz M, Hurley CK, Noreen H, Oudshoorn M, Petersdorf E, Setterholm M, Spellman S, Weisdorf D, Williams TM, Anasetti C. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. Blood. 2007 Dec 15;110(13):4576-83.PMID: 17785583.

[8] Hurley CK, Foeken L, Horowitz M, Lindberg B, McGregor M, Sacchi N; WMDA Accreditation and Regulatory Committees.. Standards, regulations and accreditation for registries involved in the worldwide exchange of hematopoietic stem cell donors and products. Bone Marrow Transplant. 2010 May;45(5):819-24.doi: 10.1038/bmt.2010.8. PMID: 20173794.

[9] Ferrara GB, Bacigalupo A, Lamparelli T, Lanino E, Delfino L, Morabito A, Parodi AM, Pera C, Pozzi S, Sormani MP, Bruzzi P, Bordo D, Bolognesi M, Bandini G, Bontadini A, Barbanti M, Frumento G. Bone marrow transplantation from unrelated donors: the impact of mismatches with substitutions at position 116 of the human leukocyte antigen class I heavy chain. Blood. 2001 Nov 15;98(10):3150-5. PMID: 11698304.

immunologicznej na prezentowany peptyd, którym był liczący 11 aminokwasów fragment antygenu BZLF wirusa Epsteina-Barra (peptyd EPLP)[10] Utworzenie kompleksu B*35:01/EPLP/TCR, jak udowodniono eksperymentalnie przez test cytotoksyczności in vitro, powodowało silną odpowiedź immunologiczną, podczas gdy kompleks B*35:08/EPLP/TCR nie wywoływał znaczącej reakcji. Jedyną różnicą pomiędzy allelami HLA-B*35:01 i B*35:08 jest pojedyncza substytucja aminokwasowa w pozycji 256 (Leu – Arg). Informację tę wykorzystano w dalszych analizach, traktując poziom energii swobodnej oryginalnych kompleksów jako punkt odniesienia. Założeniem analizy było stworzenie modeli komparatywnych (bazujących na opisanych wyżej strukturach matrycowych homologicznych sekwencji HLA) z założeniem zmiany tylko jednego elementu – cząsteczki HLA i wykonanie minimalizacji energii zmodyfikowanych kompleksów w celu obliczenia energii swobodnej kontaktu TCR-MHC i TCR-MHCp. Obliczenia zostały wykonane przy użyciu procedury wielokrotnej minimalizacji modułu MacroModel9.5 w polu OPLS 2001, dostępnego w pakiecie oprogramowania Schrodinger®, New York, NY[11]. Sekwencja prezentowanego peptydu EPLP pozostała nie zmieniona, natomiast jego konformacje do dokowania zostały wygenerowane za pomocą algorytmu LLMOD (large-scale low-mode) modułu MacroModel9.5. Uzyskano ponad 800 możliwych konformacji, z których do dalszych analiz wybrano 75, o najniższych wartościach energii i RMSD (Root-Mean-Square Deviation of atomic positions). Wybór alleli HLA do podstawienia został oparty na wynikach analizy faktycznych danych dotyczących procedur przeszczepienia komórek krwiotwórczych od dawcy niespokrewnionego pochodzących z dwóch źródeł: Centralnego Rejestru Niespokrewnionych Dawców Szpiku i Krwi Pępowinowej prowadzonego przez Poltransplant (dane dotyczące polskich pacjentów z lat 2001-2008) oraz opublikowanych danych International Histocompatibility Working Group. Łącznie dane dotyczyły 1757 przypadków transplantacji, 1089 przeszczepień wykonano od dawcy w pełni zgodnego, 668 od dawcy z niezgodnością. Najczęściej obserwowanymi niezgodnościami w analizowanym zbiorze były różnice w HLA-B: B*35:01(biorca)/B*35:03(dawca) i B*27:02(biorca)/B*27:05(dawca). Dla alleli HLA-B*35:03, B*27:02 i B*27:05 skonstruowano na podstawie matrycy 2NX5 modele struktury trójwymiarowej za pomocą zautomatyzowanej procedury modelowania dostępnej na serwerze SWISS-MODEL[12]. Na pierwszym etapie procedury struktury kompleksu TCR-pMHC dla 75 różnych konformacji EPLP po podmianie struktury HLA na model dla danego allelu podlegały minimalizacji, w następnym kroku obliczano w polu OPLS 2001 energię dla strefy kontaktu TCR-MHC (w jej skład wchodziły reszty MHC w promieniu 5Å od cząsteczki TCR i reszty TCR w promieniu 5 Å od cząsteczki MHC(HLA)) oraz strefy kontaktu TCR-p.

---

[10] Tynan FE, Reid HH, Kjer-Nielsen L, Miles JJ, Wilce MC, Kostenko L, Borg NA, Williamson NA, Beddoe T, Purcell AW, Burrows SR, McCluskey J, Rossjohn J. A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule. Nat Immunol. 2007 Mar;8(3):268-76. PMID: 17259989.

[11] http://www.schrodinger.com/

[12] Swiss Institute of Bioinformatics, Basel, Switzerland

Wyniki kalkulacji energii interakcji dla wybranych kompleksów TCR (T-cell Receptor) dawcy – pHLA biorcy wykazały, **że energia kontaktu w zakresie tzw. synapsy komórkowej różniła się przy wymianie alleli HLA w kompleksie HLA–TCR, nawet przy zachowaniu tego samego peptydu i receptora TCR, przy czym bardzo duży wpływ na energię stabilizacji kompleksu miała konformacja prezentowanego peptydu. Drugą ważną obserwacją była zależność energii interakcji TCR-MHC i TCR–peptyd od liczby i lokalizacji substytucji aminokwasowych między allelami HLA.** Różnice w całkowitej energii interakcji dla kompleksów minimalizowanych dla modeli HLA z wyższą liczbą substytucji (np. między B*27:02 a B*27:05) lub między różnymi grupami alleli (B27 i B35) były większe niż w przypadku alleli różniących się tylko jednym aminokwasem B*35:01/B*35:03. Wynik taki jest zgodny z oczekiwaniem, ale stał w sprzeczności z obserwowanymi wynikami transplantacji w grupach pacjentów przeszczepionych z niezgodnością B*35:01/B*35:03 i B*27:02/B*27:05. Wbrew oczekiwaniom krzywa przeżycia dla tej ostatniej grupy była wyraźnie lepsza niż dla grupy pierwszej. Jednak analizując wyłącznie interakcje TCR-MHC, dla ścisłej strefy kontaktu między tymi dwiema cząsteczkami zaobserwowano zupełnie odwrotne relacje. Tym razem różnice dla energii interakcji obliczone dla alleli B*35:01 i B*35:03 okazały się być wyższe niż w przypadku B*27:02 i B*27:05, mimo większej liczby różnic aminokwasowych między allelami tej drugiej pary. Jedyna różnica aminokwasowa między B*3501 i B*3503 występuje poza strefą kontaktu MHC-TCR zdefiniowaną na podstawie danych literaturowych i odległości międzycząsteczkowych mierzonych dla modelu struktury. Jest ona zlokalizowana w alfa-helisie rowka wiążącego, czyli w tzw. regionie APS (Antigen Presenting Site), uznawanym za bardzo ważny dla specyficzności cząsteczki HLA. Możliwe, że substytucje zlokalizowane w rowku wiążącym peptyd mogą wpływać na energię interakcji MHC-TCR w sposób pośredni.

Porównując całkowitą energię interakcji dla kompleksu, który nie aktywował komórek T in vivo z wartościami otrzymanymi dla oryginalnego kompleksu B*35:01-pTCR i zmodyfikowanych kompleksów z modelami B*2705 i B*2702 można wnioskować, że te ostatnie nie powinny inicjować reakcji immunologicznej.

W świetle otrzymanych wyników można stwierdzić, że energie związane z interakcją MHC-TCR na pierwszym etapie rozpoznawania kompleksu p-MHC przez receptor TCR mogą pomóc przewidzieć powstanie reakcji immunologicznej. Małe różnice w energii interakcji ($E_{MHC-TCR}$ + E_{TCR-p}) dla kompleksów powstałych z udziałem HLA dawcy i biorcy wydają się być skorelowane z dłuższym czasem przeżycia biorcy po przeszczepieniu. **Różnica w energii wiązania pomiędzy allelami HLA nie jest bezpośrednio zależna od liczby różnic aminokwasowych między nimi, a raczej od ich lokalizacji w strukturze cząsteczki.** Nawet substytucje w pozycjach odległych od powierzchni bezpośredniego kontaktu mogą wpływać na energię wiązania TCR-MHC szczególnie, jeżeli substytucja dotyczy rowka wiążącego prezentowany peptyd.

3.2. Odkrycie domeny kinazowej z atypowym miejscem aktywnym w białkach z rodziny selenoprotein (SELO), obecnych m.in. u człowieka, sugerujące ich funkcję regulatorową w odpowiedzi na stres oksydacyjny.

A Novel Protein Kinase-Like Domain in a Selenoprotein, Widespread in the Tree of Life[13].

Selenoproteiny to grupa enzymów o aktywności oksydoreduktaz, tworzące mechanizmy obrony przeciwko toksycznym substancjom ksenobiotycznym. Większość eukariotycznych genomów posiada niewielką liczbę takich białek, zwykle nie więcej niż 20, ich cechą charakterystyczną jest występowanie zmodyfikowanej reszty cysteiny, w której atom siarki jest zastąpiony przez selen. Synteza katalitycznej selenocysteiny wymaga utrzymywania kosztownego aparatu enzymatycznego, wiąże się też z zapotrzebowaniem na selen w środowisku. Biorąc pod uwagę tak wysokie koszty ewolucyjne trzeba założyć, że selenobiałka pełnią istotne funkcje w proteomie eukariotycznym. Rodzina ludzkich selenoprotein liczy 25 sekwencji, większość ma aktywność oksydoreduktazową z selenocysteiną w miejscu aktywnym, kilka z nich nie ma jeszcze przypisanych i zbadanych funkcji. Jednym z takich białek jest ludzka selenoproteina O (SELO), od kilku lat typowana jako jedno z potencjalnie najciekawszych białek o niepoznanej jeszcze funkcji[14], która stała się przedmiotem przedstawionej poniżej analizy bioinformatycznej.

Pierwszym etapem analizy było zastosowanie narzędzi do poszukiwania odległych homologii i przewidywania struktury trzeciorzędowej: FFAS03 oraz HHpred, które wykazały dla centralnego regionu sekwencji SELO podobieństwo do domen kinazowych (m.in. APH, Kdo/WaaP oraz innych rodzin należących do kinazo-podobnego klanu Pfam CL0126). Dla pozostałych fragmentów sekwencji nie uzyskano istotnych predykcji. Następnym krokiem było poszukiwanie w dostępnych zasobach wszystkich możliwych homologów kinazo-podobnej domeny SELO i konstrukcja drzewa filogenetycznego zawierającego jak najwięcej sekwencji pochodzących z odległych organizmów w celu zbadania jej rozpowszechnienia oraz konserwacji motywów w rejonie miejsca aktywnego. Wykorzystano w tym celu narzędzia do poszukiwania odległych homologii o zwiększonej czułości w porównaniu do klasycznych algorytmów: tzw. Saturated BLAST i HHsenser. Kaskadowe wyszukania, wykorzystujące zbiory reprezentatywnych trafień do inicjacji kolejnych iteracji, dały w rezultacie zbiór kilkuset przypuszczalnych homologów. Zostały one dodatkowo zweryfikowane standardowymi narzędziami w celu potwierdzenia podobieństwa do rodziny Pfam UPF0061, do której przypisano ludzką sekwencję

---

[13] Dudkiewicz M, Szczepińska T, Grynberg M, Pawłowski K. PLoS One. 2012;7(2):e32138. doi: 10.1371/journal.pone.0032138. Epub 2012 Feb 16,PMID: 22359664.

[14] Galperin MY, Koonin EV. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Res. 2004 Oct 12;32(18):5452-63.Review. PMID: 15479782.

SELO. Ze zbioru ponad 800 sekwencji odrzucono sekwencje wykazujące ponad 70% identyczność, w celu wyboru puli sekwencji reprezentatywnych. Ostatecznie uzyskano 143 sekwencje, pochodzące z genomów bakteryjnych, archebakteryjnych, eukariotycznych oraz przypisanych do metagenomu morskiego (pochodzące z próbek środowiskowych pobranych z Morza Sargassowego). Sekwencje te zostały wykorzystane do konstrukcji wielokrotnego dopasowania i drzewa ML (maximum likelihood) z uwzględnieniem cech taksonomicznych oraz konserwacji występujących przy C-końcu sekwencji aminokwasowej motywów CxxC(U)> zawierających potencjalnie selenocysteinę o aktywności oksydoreduktazowej. Do zobrazowania podobieństw między typowymi rodzinami kinazowymi a rodziną SELO (UPF0061) wykorzystano algorytm CLANS. Uzyskany graf wskazywał na silne związki między SELO a klasycznymi rodzinami kinaz serynowych i tyrozynowych, ale także licznymi rodzinami pseudokinazowymi.

Analiza drzewa filogenetycznego dla homologów SELO i rozpowszechnienia tej domeny wśród różnych grup taksonomicznych pozwoliła na sformułowanie hipotezy, że domena SELO, występująca jako gen mitochondrialny u wszystkich posiadających go eukariontów i rozpowszechnionych w większości najważniejszych grup bakterii, powstała u ostatniego wspólnego przodka bakterii i eukariotycznych endosymbiontów, u nielicznych archebakterii pojawiła się prawdopodobnie drogą transferu horyzontalnego, a niektóre linie eukariotyczne utraciły ją wtórnie.

Aby zweryfikować wskazania metaserwerów o prawdopodobnej obecności domeny kinazowej lub kinazopodobnej w sekwencji SELO wykonano model struktury trójwymiarowej centralnej domeny SELO w oparciu o matrycę wskazaną przez algorytm FFAS03, którą była struktura podjednostki katalitycznej c-AMP zależnej kinazy białkowej (1CDK) w kompleksie z analogiem ATP i jonami manganu. Modelowanie poprzedzono dokładną analizą dopasowania sekwencji do matrycy wykonanego za pomocą metody profilowej z manualnymi poprawkami wprowadzonymi na podstawie wyników przewidywania struktur drugorzędowych (wykorzystano konsensus dla wskazań algorytmów Jpred, SOPMA i PsiPred). Model został wykonany przy pomocy zmodyfikowanych skryptów pakietu Modeller9v8[15], z podstawień jonów Mn^{2+} przez Mg^{2+} oraz ANP przez ATP. Uzyskane wyniki sugerowały, że prawdopodobna domena kinazowa SELO składa się z mniejszej podjednostki N-końcowej zbudowanej głównie z elementów struktury beta, wiążącej typowy dla kinaz ligand-ATP oraz z większej, głównie alfa-helikalnej podjednostki C-końcowej. Kolejność elementów beta i alfa zgadza się mniej więcej z typową strukturą rdzenia domeny kinazowej. Z dziewięciu typowych dla PKA subregionów u SELO udało się zidentyfikować sześć, w tym m.in. wiążącą ATP pętlę glicynową, pętlę katalityczną oraz motyw DFG, miejsce wiązania jonu magnezu. Z krytycznych dla aktywności kinazowej reszt udało się potwierdzić zachowanie reszt: Lys72 zaangażowanej w wiązanie grup fosforanowych cząsteczki ATP, Glu91,

[15] Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993 Dec 5;234(3):779-815. PMID: 8254673.

stabilizującej Lys72 przez tworzenie mostka solnego oraz Asp184, odpowiedzialnej za wiązanie jonu magnezu. Motyw HRD/YRD charakterystyczny dla miejsca aktywnego klasycznych kinaz nie jest zachowany w domenie SELO, zastępuje go motyw HGV, brakuje więc potencjalnej katalitycznej reszty Asp166, jak jednak wskazują dane literaturowe brak Asp166 nie oznacza braku aktywności kinazowej[16]. **Fakt, że udało się skonstruować i zwalidować model komparatywny dla sekwencji SELO w oparciu o matrycę typowej struktury kinazowej, przemawia za hipotezą o obecności domeny kinazopodobnej w tym białku.**

Ostatnim etapem badania była analiza dostępnych danych dotyczących ekspresji genów z rodziny SELO oraz analiza sąsiedztwa genomicznego. W atlasie ekspresji tkankowej BiopGPS można zauważyć podwyższoną ekspresję ludzkich i mysich genów SELO w tkankach wątroby. W przypadku homologów drożdżowych i bakteryjnych zaobserwowano podwyższony poziom transkryptów SELO w warunkach tlenowych w porównaniu z beztlenowymi, co może wskazywać na rolę tego białka w reagowaniu na stres oksydacyjny. Także wyniki badania otoczenia genomicznego bakteryjnych homologów SELO wskazują na taką właśnie funkcję – w otoczeniu genu znajdują się liczne loci zaangażowane w reakcję na różnego rodzaju stres komórkowy.

Analiza sekwencji z wykorzystaniem metod profilowych pozwoliła także na wykrycie odległej homologii między SELO a grupą bakteryjnych białek mchC zaangażowanych w syntezę, maturację i eksport bakteriobójczych antybiotyków - mikrocyn. Produkcja i wydzielanie tzw. antybiotyków cyklu redoks jest jednym z mechanizmów obrony przed stresem oksydacyjnym.

Przeprowadzone analizy pozwoliły na wysunięcie i umotywowanie przypuszczenia, że rodzina białek, do której należy ludzka selenoproteina O, to ewolucyjnie stare oksydoreduktazy z prawdopodobną aktywnością enzymatyczną, której nie udało się jednak zidentyfikować w sposób jednoznaczny, zaangażowane w kaskadę sygnałową powstającą w odpowiedzi na stres oksydacyjny.

Rodziną selenobiałek SELO zainteresował się zespół Prof. Tagliabracci z University of Texas Southwestern Medical Center w Dallas. Dzięki szeroko zakrojonym badaniom eksperymentalnym, włączając rozwiązanie struktury trójwymiarowej bakteryjnego homologa SELO metodą krystalografii rentgenowskiej, prowadzonym we współpracy z IBB PAN i Katedrą Bioinformatyki i Doświadczalnictwa SGGW udało się opublikować w 2018 artykuł w *Cell* donoszący o odkryciu nietypowej aktywności enzymatycznej u białek SELO- okazały się one być ampylazami, nie kinazami, zachowując jednocześnie przewidziany przez nas rdzeń struktury typowy dla kinaz i wiążąc ATP w przewidzianym miejscu

[16] Shi F, Telesco SE, Liu Y, Radhakrishnan R, Lemmon MA. ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. Proc Natl Acad Sci U S A. 2010 Apr 27;107(17):7692-7. PMID:20351256.

aktywnym, jednak w odwróconej o 180^0 konformacji[17]. **Selenoproteiny okazały się być bardzo nietypowymi pseudokinazami, z aktywnością enzymatyczną, której do tej pory nie raportowano dla żadnego białka z szeroko definiowanej rodziny kinaz i kinase-like. Ampylacja, obok fosforylacji jest jeszcze jednym rozpoznanym mechanizmem komórkowego szlaku przekazywania sygnałów, być może jej znaczenie jest większe niż do tej pory uważano**. Co więcej, zespołowi z UTSMC w Dallas udało się dowieść eksperymentalnie, że SELO ampyluje glutaredoksynę, zależny od glutationu enzym biorący udział w regulacji równowagi redoks w komórce i dzięki temu mechanizmowi bierze udział w regulacji S-glutationylacji białek i odpowiedzi na stres oksydacyjny.

[17] Sreelatha A, Yee SS, Lopez VA, et al. Protein AMPylation by an Evolutionarily Conserved Pseudokinase. Cell. 2018 Oct 18;175(3):809-821.e19. doi: 10.1016/j.cell.2018.08.046. Epub 2018 Sep 27. PubMed PMID: 30270044.

3.3. Odkrycie nowej rodziny kinaz regulowanych jonami wapnia, Secretory Pathway Kinases (SPK), prawdopodobnie związanej z rozwojem schorzeń z grupy zaburzeń neurologicznych

A Novel Predicted Calcium-Regulated Kinase Family Implicated in Neurological Disorders[18].

Kinazy białkowe, najważniejsze cząsteczki efektorowe szlaków sygnałowych w komórce uruchamiające kaskady sygnałów przez fosforylację kolejnych substratów białkowych są jedną z najlepiej opisanych i najchętniej badanych rodzin białek, szczególnie w proteomie ludzkim, ze względu na ich szczególną rolę regulatorową i potencjalne zastosowanie w medycynie i przemyśle farmaceutycznym. Natrafienie na nieopisaną rodzinę kinaz, szczególnie zawierającą sekwencje występujące u człowieka jest mało prawdopodobne, a jednak udało nam się metodami bioinformatycznymi wykryć domenę kinazową w nieadnotowanej rodzinie białkowej FAM69. Wyniki analizy modeli struktur trójwymiarowych i konserwacji motywów katalitycznych charakterystycznych dla domen kinazowych obecnych u czterech z pięciu ludzkich białek FAM69 wskazują, że mogą one mieć aktywność fosfotransferazową. Obecność wiążącej jony Ca^{2+} tzw. domeny EF-hand w obrębie domeny kinazowej u niektórych reprezentantów rodziny sugeruje, że mogą one funkcjonować jako kinazy regulowane jonami wapnia. Geny FAM69 są w większości niescharakteryzowane molekularnie, jednak doniesienia literaturowe wskazują na ich związek z zaburzeniami neurologicznymi, np. delecja genu C3ORF58(DIA1) została stwierdzona u chorych na autyzm.

Pierwszym etapem analizy było zastosowanie metod profilowych do wyszukania odległych homologii dla reprezentanta nieopisanej rodziny Pfam PIP49_C, ludzkiego białka FAM69A. Zarówno wskazania FFAS03, jak i HHpred sugerowały podobieństwo do klasycznych kinaz serynowo-treoninowych i tyrozynowych. Obszar istotnej homologii ograniczał się do regionu 225-415 sekwencji FAM69A. W analizie zastosowano również algorytm CLANS na sekwencjach reprezentatywnych 17 rodzin należących do szeroko pojętego klanu białek kinazo-podobnych. Uzyskany graf wskazał na istotne związki PIP49_C ze wszystkimi badanymi rodzinami na wszystkich 4 badanych poziomach istotności (od 0,1 do 1E-10). W kolejnym kroku skonstruowano wielokrotne dopasowania najpierw dla podrodzin PIP49_C : FAM69A, B, C oraz DIA1 i DIA1R za pomocą algorytmu MUSCLE, a następnie zestawiono dopasowania dla podrodzin z uwzględnieniem więzów dla wybranych motywów kinazowych za pomocą programu Promals3D i przyrównano do sekwencji wzorcowych kinaz o znanej strukturze, uzyskując spójny MSA, do którego dodano przewidywania struktury drugorzędowej dla przedstawicieli rodzin (uzyskane za pomocą serwerów TMHMM, Phobius, Jpred i PsiPred). Model strukturalny ludzkiego białka FAM68A został opracowany metodą komparatywną. Matryce do modelowania wybrano na podstawie wskazań serwera FFAS03, bazującego na porównaniu profili, uzupełnionych o wyniki HHPred (porównanie

[18] Dudkiewicz M, Lenart A, Pawłowski K. PLoS One. 2013 Jun28;8(6):e66427, PMID:23840464.

modeli HMM). Najlepsze wyniki dawały dwie struktury kinaz zdeponowane w bazie PDB (3HGK i 3S95) dla białka ludzkiego i roślinnego, oraz, z uwagi na obecność nietypowej pętli wiążącej wapń (EF-hand) struktura 2PMY (domena EF ludzkiego białka RASEF). Ostatecznie do modelowania wykorzystano połączenie trzech matryc, a do odtworzenia pozycji wiązania ligandu wykorzystano dodatkowo strukturę kinazy A (1CDK), wykrystalizowaną razem z cząsteczką ANP i jonem Mn^{2+}. Dwie matryce kinazowe zostały dopasowane do siebie za pomocą algorytmu FATCAT[19]. Do dopracowania MSA między matrycami a modelowaną sekwencją wykorzystano kombinację pojedynczych dopasowań uzyskanych metodami FATCAT, HHpred i FFAS03 oraz niewielkie ręczne poprawki uwzględniające dopasowanie struktur drugorzędowych. Model wykonano za pomocą zautomatyzowanej procedury Modeller9v8, a jego walidacja została przeprowadzona za pomocą serwera MetaMQAP[20].

Wyniki analizy sekwencji i modelu przestrzennego wykazały, że FAM69A wykazuje wysokie podobieństwo do strukturalnej podjednostki C-końcowej klasycznej kinazy. Podobieństwo w regionie N-końcowym uzyskiwało wartości na granicy istotności, jednak po usunięciu insercji - fragmentu sekwencji zawierającego motyw EF-hand, reszta sekwencji wykazała istotne ogólne podobieństwo do kinaz białkowych. Być może właśnie ta insercja przeszkadzała w automatycznym odkryciu homologii w procesie automatycznej adnotacji genomu ludzkiego.

Jeżeli chodzi o zachowanie typowych motywów i reszt katalitycznych, zarówno w dopasowaniu sekwencji FAM69 jak i w modelu struktury można zaobserwować konserwację reszty Asp odpowiadającej katalitycznej D166 klasycznej kinazy PKA. Jedynym wyjątkiem jest CXORF36, przypuszczalna pseudokinazą. Obecne są także odpowiedniki reszt N171 i D184 PKA odpowiedzialne za wiązanie jonów magnezu oraz odpowiedniki reszt lizyny i kwasu glutaminowego w regionie β-3 i helisie α-C PKA zaangażowane w wiązanie grup fosforanowych ATP. Z drugiej strony sekwencja FAM69A zdaje się nie posiadać typowego motywu wiązania ATP GxGxxG, co zdaje się wskazywać na nietypowy sposób wiązania ligandu w tej grupie białek. Następnym ciekawym spostrzeżeniem jest podwyższona zawartość cysteiny w sekwencjach FAM69 (potwierdzona analizą statystyczną dla zawartości cysteiny dla wszystkich znanych rodzin kinazowych). Większość z cystein jest zachowana i może uczestniczyć w tworzeniu mostków dwusiarczkowych, co sugeruje też model struktury trójwymiarowej. Jeden z nich , między resztami Cys293 i Cys331, blisko przypuszczalnie katalitycznej reszty Asp, stabilizuje prawdopodobnie konformację miejsca aktywnego.

Zgodnie z przewidywaniami algorytmów rozpoznających domeny transmembranowe, FAM69A, B i C posiadają domeny błonowe, podczas gdy C3ORF58 i CXORF36 wykazują raczej obecność sekwencji

---

19 Ye Y, Godzik A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics. 2003 Oct;19 Suppl 2:ii246-55. PMID:14534198.

20 Pawlowski M, Gajda MJ, Matlak R, Bujnicki JM. MetaMQAP: a meta-server for the quality assessment of protein models. BMC Bioinformatics. 2008 Sep 29;9:403. PMID: 18823532.

sygnałowych. Doniesienia literaturowe wskazują z kolei na lokalizację białek FAM69 w retikulum endoplazmatycznym[21], natomiast obecność produktu genu C3ORF58 potwierdzono w aparacie Golgiego[22]. Analiza dostępnych w bazach danych ekspresyjnych wykazała, że ekspresja FAM69A nie jest specyficzna tkankowo, podczas gdy transkrypty FAM69B i C występują głównie w tkance mózgu i oka. Ekspresja genu C3ORF58 została zaobserwowana w tkance chrząstki, szczególnie w rozwijających się chondrocytach, jak również w zawiązkach zębowych. Dane literaturowe są zgodne z danymi ekspresyjnymi, sugerując związek niektórych genów z rodziny FAM69 z licznymi zaburzeniami neuronalnymi u człowieka. Jedna z dwóch największych delecji chromosomowych towarzyszących autyzmowi obejmuje region kodujący C3ORF58(DIA1), dodatkowo wiadomo, że ekspresja tego genu jest regulowana przez aktywność neuronalną[23]. Z kolei CXORF36 (DIA1R) jest łączony z zespołem łamliwego chromosomu X (FXS), którego mechanizm molekularny jest prawdopodobnie bardzo zbliżony do tego wywołującego autyzm, ponieważ 30% pacjentów ze zdiagnozowanym FXS wykazuje także symptomy autyzmu[24]. Niektóre publikacje łączą także region chromosomu zawierający locus CXORF36 z zaburzeniami neurologicznymi, zwiększoną skłonnością do autyzmu u mężczyzn i zespołem Kabuki, związanym z wrodzonym upośledzeniem umysłowym[25].

Uzyskane wyniki analiz bioinformatycznych i poszukiwań literaturowych skłaniają do przedstawienia hipotezy, że białka FAM69 są ważnymi kinazami regulującymi prawdopodobnie procesy transportu z aparatu Golgiego do błony komórkowej lub z ER do aparatu Golgiego, być może przez fosforylację białek ulegających sekrecji. Jak wykazano prawidłowe funkcjonowanie transportu pęcherzykowego jest krytyczne dla rozwoju neuronów u człowieka, a wystąpienie nieprawidłowości może doprowadzić do opóźnienia umysłowego[26]. Być może wspólnym

---

[21] Tennant-Eyles AJ, Moffitt H, Whitehouse CA, Roberts RG. Characterisation ofthe FAM69 family of cysteine-rich endoplasmic reticulum proteins. Biochem Biophys Res Commun. 2011 Mar 18;406(3):471-7. PMID: 21334309.

[22] Takatalo M, Järvinen E, Laitinen S, Thesleff I, Rönnholm R. Expression of the novel Golgi protein GoPro49 is developmentally regulated during mesenchymal differentiation. Dev Dyn. 2008 Aug;237(8):2243-55. PMID: 18651652.

[23] Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS, Mukaddes NM, Balkhy S, Gascon G, Hashmi A, Al-Saad S, Ware J, Joseph RM, Greenblatt R, Gleason D, ErteltJA, Apse KA, Bodell A, Partlow JN, Barry B, Yao H, Markianos K, Ferland RJ,Greenberg ME, Walsh CA. Identifying autism loci and genes by tracing recent shared ancestry. Science. 2008 Jul 11;321(5886):218-23. PMID: 18621663.

[24] Hagerman R, Hoem G, Hagerman P. Fragile X and autism: Intertwined at themolecular level leading to targeted treatments. Mol Autism. 2010 Sep 21;1(1):12. PMID: 20858229.

[25] Lederer D, Grisart B, Digilio MC, Benoit V, Crespin M, Ghariani SC, Maystadt I, Dallapiccola B, Verellen-Dumoulin C. Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. Am J Hum Genet.2012 Jan 13;90(1):119-24. PMID: 22197486.

[26] Giannandrea M, Bianchi V et a. Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy and macrocephaly. Am J Hum Genet. 2010 Feb 12;86(2):185-95. PMID: 20159109.

mianownikiem dla zaburzeń neurologicznych wiązanych z genami FAM69 mogą być zaburzenia komórkowych szlaków wydzielniczych w neuronach.

3.4. Odkrycie bakteryjnych aneksyn prawdopodobnych przodków aneksyn eukariotycznych.

Bioinformatics Analysis of Bacterial Annexins – Putative Ancestral Relatives of Eukaryotic Annexins[27].

Aneksyny to rodzina białek rozpowszechnionych wśród organizmów eukariotycznych, oddziałujących z błonami biologicznymi, mających zdolność odwracalnego wiązania jonów Ca^{2+}. Aneksyny mogą pełnić bardzo różne funkcje w komórce, ale za ich główną rolę uważa się udział w transporcie pęcherzykowym, zwłaszcza w endocytozie. Zaangażowanie w tak podstawowe procesy tłumaczy zasięg ich występowania. Struktura aneksyn jest bardzo charakterystyczna, wzorcowe aneksyny kręgowców są złożone z czterech homologicznych domen (tzw. powtórzeń aneksynowych – annexin repeats) o długości ok. 70 aminokwasów. Rdzeń domeny aneksynowej jest odpowiedzialny za odwracalne, zależne od stężenia jonów wapnia, wiązanie cząsteczek fosfolipidów błonowych. Typowa cząsteczka aneksyny eukariotycznej składa się z czterech powtórzeń, chociaż zdarzają się także aneksyny 8-rdzeniowe. Powszechnie przyjęta hipoteza na temat pochodzenia aneksyn zakłada zajście podwójnej duplikacji fragmentu genomu. Niedawne odkrycie jednodomenowej aneksyny bakteryjnej u *Cytophaga hutchinsonii* zdaje się potwierdzać tę hipotezę[28]. **W celu statystycznej analizy zjawiska i znalezienia dodatkowych dowodów na poparcie tezy o bakteryjnych korzeniach aneksyn wykonaliśmy szereg analiz bioinformatycznych, starając się za pomocą narzędzi do badania odległych homologii zidentyfikować i zbadać wszystkie prokariotyczne homologi aneksyn.** Zastosowana procedura pozwoliła na wykrycie około 30 potencjalnych jedno- i wielodomenowych aneksyn u 17 gatunków bakterii. **Sekwencje bakteryjnych odpowiedników aneksyn wykazywały odległe, ale statystycznie istotne podobieństwo do profili białkowych skonstruowanych dla aneksyn eukariotycznych. Niektóre aneksyny bakteryjne były wyposażone w dodatkowe atypowe domeny, np. charakterystyczne dla toksyn.** Zróżnicowanie sekwencji i architektury domenowej bakteryjnych aneksyn, o wiele większe niż u ich eukariotycznych homologów, przemawia za teorią o ich wcześniejszej genezie i dłuższej historii ewolucyjnej. Przeciwko hipotezie o bakteryjnym przodku większości współczesnych aneksyn świadczy małe rozpowszechnienie aneksyno-podobnych sekwencji wśród linii bakteryjnych.

---

[27] Kodavali PK, Dudkiewicz M, Pikuła S, Pawłowski K. PLoS One. 2014 Jan 16;9(1):e85428. doi: 10.1371/journal.pone.0085428,PubMed PMID: 24454864.

[28] Morgan RO, Martin-Almedina S, Garcia M, Jhoncon-Kooyip J, Fernandez MP. Deciphering function and mechanism of calcium-binding proteins from their evolutionary imprints. Biochim Biophys Acta. 2006 Nov;1763(11):1238-49. PMID: 17092580.

3.5. Przypisanie nieadnotowanej rodziny białkowej DUF2362 do klanu Macro – odkrycie potencjalnego nowego uczestnika komórkowego szlaku sygnałowego opartego na ADP-rybozylacji.

C12orf4, a novel conserved family of macro-like domains, likely new players in ADP-ribosylation signaling[29]

Ludzkie białko C12ORF4 należy do rodziny DUF2362, obejmującej domeny o nieznanej funkcji obecne i zachowane w wielu liniach Eukaryota, głównie Metazoa. Jedyna dostępna informacja o przypuszczalnej funkcji C12ORF4 to jego powiązanie z degranulacją komórek tucznych układu odpornościowego oraz z genetycznymi przyczynami dziedzicznej autosomalnej niepełnosprawności intelektualnej. [30]

Wykonane przez nas analizy bioinformatyczne dostarczyły silnych dowodów na poparcie tezy, że białko to jest nowym członkiem ważnego białkowego klanu Makro, łączącego białka posiadające makrodomeny, charakterystyczne elementy strukturalne odpowiedzialne za wiązanie reszt ADP-rybozy.

Analiza podobieństwa sekwencji białek zaliczanych do rodziny DUF2362 w stosunku do innych przedstawicieli nadrodziny Makro, wykonana zarówno standardowymi metodami przyrównania sekwencji (BLAST, grupowanie za pomocą algorytmu CLANS), jak i metodami profilowymi, pozwoliła na sformułowanie tezy o ich homologii sekwencyjnej, a także prawdopodobnie strukturalnej i funkcjonalnej z makrodomenami. Narzędzia poszukiwania odległych homologii wskazywały na silny, pojedynczy sygnał typowy dla profilu makrodomeny w C-końcowej części sekwencji ludzkiego białka C12ORF4 (rejon 322-552). Największe podobieństwo stwierdzono w przypadku domeny makro typu "eraser", mającej właściwości nie tylko wiążące ADP-rybozę, ale i katalityczne – zdolność usuwania dodanych post-translacyjnie do cząsteczki białka reszt ADP-rybozy, a ponadto występującej zwykle tylko w jednej kopii (w przypadku makrodomen wiążących, typu „reader" obecne są najczęściej 3 powtórzenia motywu domeny makro). Na matrycy takiej właśnie makrodomeny typu "eraser" pochodzącej z ludzkiego MacroD2 udało się zbudować model homologiczny przypuszczalnej struktury trójwymiarowej białka C12ORF4, z użyciem elementów konstrukcji pętli ab initio oraz dokowania przypuszczalnego ligandu – ADP-rybozy. Do predykcji lokalizacji miejsca wiązania ADP-rybozy w analizowanym modelu użyto metody skanowania bazy danych PDB w poszukiwaniu struktur podobnych i zawierających ligandy. Udało się potwierdzić obecność i zachowanie w całej rodzinie DUF2363 kluczowych dla MacroD2 reszt katalitycznych z pętli 1 Asp i Asn, sekwencja pętli 2 nie wykazywała jednak wyraźniej homologii ani do wzorca dla makrodomen typu: eraser", ani „reader".

---

[29] Dudkiewicz M, Pawłowski K. PeerJ. 2019 May 1;7:e6863. doi: 10.7717/peerj.6863,PMID: 31106069.

[30] Alazami AM, Patel N, et al. 2015. Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. Cell Reports 10(2):148–161. PMID 25558065

Wcześniejsze doniesienia literaturowe 31, wraz z naszymi odkryciami, które łączą C12ORF4 z ADP-rybozylacją, rzucają nowe interesujące światło na znaną funkcję PARP-1 w degranulacji, a konkretnie w modulowaniu wytwarzania cytokin związanych z astmą32. Inhibitory PARP1 i PARP14 są już brane pod uwagę w leczeniu astmy i na razie testowane na modelach zwierzęcych33. Zaobserwowano m.in., że ekspozycja na alergeny wywoływała u zwierząt uczulonych gwałtowny wzrost aktywności PARP, zwężenie oskrzeli, inflację przestrzeni powietrznej płuc i degranulacja komórek tucznych, natomiast inhibicja PARP zapobiegała powyższym zmianom morfologicznym i biochemicznym w płucach34

Nasze wyniki pozwalają nam sformułować hipotezę, że posiadające makrodomenę białko C12ORF4, jako domniemana „gumka" ADPr, reguluje zależny od PARP szlak sygnałowy ADP-rybozylacji białek, podobnie jak PARG (glikohydrolaza Poli (ADP-rybozy)), której antagonistyczna do PARP rola w chorobach dróg oddechowych była już wcześniej znana. Fakt, że ekspresja C12ORF4 może wpływać na fosforylację kilku kinaz umożliwia spekulacje, że wpływ PARP i C12ORF4 na degranulację może być bardzo bezpośredni i odbywać się przez ADP-rybozylację kinaz. Znane są przypadki regulacji ścieżki sygnalizacji kinazowej na drodze bezpośredniej ADP-rybozylację kluczowych kinaz przez efektory bakteryjne, a także przez endogenne ludzkie ADP-rybozylotransferazy.

Interesującym powiązaniem między przypisanym C12ORF4 fenotypem związanym z degranulacją a jego połączeniem z niepełnosprawnością umysłową jest fakt, że degranulacja komórek tucznych jest jednym z procesów zaangażowanych w zapalenie mózgu[35] [36]. Z kolei zapalenie mózgu może prowadzić do ID i powiązanych komplikacji.

Podsumowując, zgromadzono mocne dowody in silico, że rodzinę DUF2362 należy zaliczyć do klanu makrodomen. Podobieństwo do znanych domen Makro jest jasne i sugestywne. Pozostaje niejasne, czy ta nowa domena Makro ma aktywność katalityczną, czy tylko funkcję „czytnika" sygnału ADP-rybozylacji, jednak kilka argumentów przemawia za funkcją katalityczną, a nie

[31] Mazuc E, Guglielmi L, et al.. 2014. In-cell intrabody selection from a diverse human library identifies C12orf4 protein as a new player in rodent mast cell degranulation. PLOS ONE 9(8)e104998. PMID25122211

[32] Sethi GS, Dharwal V, Naura AS. Poly(ADP-Ribose)Polymerase-1 in Lung Inflammatory Disorders: A Review. Front Immunol. 2017 Sep 19;8:1172 PMID: 28974953.

[33] Zaffini R, Gotte G, Menegazzi M. Asthma and poly(ADP-ribose) polymerase inhibition: a new therapeutic approach. Drug Des Devel Ther. 2018 Feb 12;12:281-293. PMID: 29483769.

[34] Suzuki Y, Masini E, Mazzocca C, Cuzzocrea S, Ciampa A, Suzuki H, Bani D. Inhibition of poly(ADP-ribose) polymerase prevents allergen-induced asthma-like reaction in sensitized Guinea pigs. Journal of Pharmacology and Experimental Therapeutics. 2004;311(3):1241–1248, PMID: 15254147

[35] Dong H, Zhang X, Wang Y, Zhou X, Qian Y, Zhang S. Suppression of brain mast cells degranulation inhibits microglial activation and central nervous system inflammation. Molecular Neurobiology. 2017;54(2):997–1007,PMID: 26797518

[36] Skaper SD, Facci L, Zusso M, Giusti P. An inflammation-centric view of neurological disease: beyond the neuron. Frontiers in Cellular Neuroscience. 2018;12:72, PMID: 29618972

pseudoenzymatyczną C12ORF4, w tym zachowanie domniemanego miejsca aktywnego, podobnego do motywów pętli katalitycznej sekwencji MacroD2 – „erasera" usuwającego stemple ADP-rybozy z substratów białkowych.

3.6. Potwierdzona laboratoryjnie identyfikacja nowego członka rodziny kinaz NFK3 – PEAK3/C19orf35: pseudokinazy inhibującej crkII. z funkcją regulatora procesu formowania się cytoszkieletu, związanej z procesami onkogenezy.

PEAK3/C19orf35 pseudokinase, a new NFK3 kinase family member, inhibits CrkII through dimerization[37].

Członkowie New Kinase Family 3 (NKF3), pseudokinazy PEAK1/SgK269 i Pragmin/SgK223 okazały się ważnymi regulatorami ruchliwości komórek i progresji onkogenezy[38],[39]. Migracja komórek nowotworowych związana jest z ich ruchliwością, wraz ze wzrostem guza rozpoczynają one ekspansję na otaczające go tkanki. Wejście komórek rakowych w stan metastazy jest zwykle przyczyną drastycznego pogorszenia rokowań dla pacjenta z nowotworem.

W powyższej publikacji dowiedziono, że C19orf35 (PEAK3), nowo zidentyfikowany członek tej rodziny jest kinazopodobnym białkiem zachowanym ewolucyjnie w linii ssaków i ptaków oraz pełni funkcję regulatora ruchliwości komórek opartej na procesach agregowania się mikrofilamentów. Ślady homologii między C19orf35 a pseudokinazami PEAK1/Sgk269 i Pragmin/SgK223 zostały zidentyfikowane metodami bioinformatycznymi, sekwencja uniknęła prawdopodobnie adnotacji jako członek rodziny NKF3 z powodu wysokiej zawartości regionów o niskiej złożoności (LCR) w swojej domenie pseudokinazowej. Pomimo tej różnicy PEAK3 zachowuje cechy charakterystyczne dla rodziny NKF3. W sekwencji zidentyfikowano reszty zdefiniowane jako triada inhibująca, które zasłaniają kinazową kieszeń wiążącą ATP[40]. Wraz z mutacjami kilku reszt katalitycznych w domenie kinazowej cechy te definiują PEAK3 jako pseudokinazę. Na podstawie analizy sekwencji i wyników mutagenezy wykazano również, PEAK3 ulega autoregulacji przez konserwatywną domenę SHED, która flankuje domenę pseudokinaz, podobnie jak ma to miejsce w białkach Sgk269 i Sgk223. Obecność domeny SHED i jej rola w oligomeryzacji definiują unikalne cechy kinaz należących do rodziny NKF3. Analiza filogenetyczna rodziny wykazała, że przodek sekwencji NKF3, który prawdopodobnie pojawił się u przodka Metazoa, był najbardziej podobny do PEAK1 i był już pseudokinazą. Białko to miało motyw NFS zamiast kanonicznego dla kinaz DFG, chociaż zachowało motyw HxD, który przekształcił się w HCD w ludzkim PEAK1, a w niektórych homologach wrócił do kanonicznej kinazowej postaci HRD. W niektórych ptasich homologach PEAK1 i Pragmin udało się natomiast zauważyć powrót do pierwotnego kinazowego

---

[37] Lopez ML, Lo M, Kung JE, Dudkiewicz M, Jang GM, Von Dollen J, Johnson JR, Krogan NJ, Pawłowski K, Jura N. Proc Natl Acad Sci U S A. 2019 Jul 30;116(31):15495-15504. doi: 10.1073/pnas.1906360116,PMID: 31311869.

[38] Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002 Dec 6;298(5600):1912-34, PMID: 12471243.

[39] Kung JE, Jura N. Structural Basis for the Non-catalytic Functions of ProteinKinases. Structure. 2016 Jan,PMID: 26745528.

[40] Ha BH, Boggon TJ. The crystal structure of pseudokinase PEAK1 (Sugen kinase 269) reveals an unusual catalytic cleft and a novel mode of kinase fold dimerization. J Biol Chem. 2018 Feb 2;293(5):1642-1650. doi:10.1074/jbc.RA117.000751,PMID: 29212708.

motywu DFG, było to jednak sprzężone z mutacją drugiego kanonicznego motywu (HxD) w kierunku zupełnego braku aktywności. **Obserwacje te wspierały hipotezę, że NKF3 jest przykładem rodziny białek kinazopodobnych, które ewoluowały jako pseudokinazy, a ich obecna funkcja w komórce wymaga braku typowej aktywności enzymatycznej.** Wyniki analizy bioinformatycznej wsparte modelowaniem struktury trójwymiarowej dimeru C12orf35 i mapowaniem konserwacji reszt dla grupy homologów PEAK3 na uzyskany model zachęciły zespół Natalii Jury z University of California San Francisco do eksperymentalnego poszukiwania funkcji dla nowego członka rodziny NKF3. Laboratoryjny screening w poszukiwaniu białek wiążących się z C19orf35 wykazał interakcje z regulatorami ruchliwości komórek, w tym z białkiem adaptorowym CrkII, produktem znanego protoonkogenu c-Crk. **Udało się wykazać, że PEAK3 wpływa na CrkII w odwrotny sposób od pozostałych członków rodziny NKF3 – hamuje działanie CrKII, podczas gdy PEAK1 i Pragmin wzmacniają jego funkcję i przekazywany sygnał komórkowy.** Mechanizm leżący u podstaw takiego antagonistycznego działania pozostaje niewyjaśniony, ale ma niewątpliwie związek z różnicami obserwowanymi w sekwencji domeny pseudokinazowej u PEAK3 i innych członków rodziny NKF3, a zwłaszcza brakiem regionu N-końcowego w sekwencji C12orf35, który może być odpowiedzialny za wzmacnianie sygnału kaskady CrkII przez białka PEAK1 i Pragmin. Wykazano również doświadczalnie, że dimeryzacja PEAK3 oparta na domenie SHED jest niezbędna do jego wiązania się z CrkII. **Perspektywa regulacji dimeryzacji PEAK3 poprzez zmiany konformacyjne w domenie pseudokinazowej daje ekscytującą możliwość farmakologicznej modulacji aktywności białka PEAK3, które zdaje się mieć obiecujący potencjał terapeutyczny[41]. Mimo że brak obecnie doniesień o związanych z nowotworami mutacjach w PEAK3, ekspresja mRNA dla genu C12orf35 jest znacząco podwyższona u pacjentów z ostrą białaczką szpikową (AML) w stosunku do grup kontrolnych.**

Wiadomo, że białka rodziny Crk odgrywają kluczową rolę w procesach inwazji i migracji komórek nowotworowych poprzez integrację i wzmocnienie sygnałów pozakomórkowych. Genetyczny knock-out CrkII wyraźnie zmniejsza zdolność do migracji i obniża złośliwy potencjał wielu ludzkich komórek rakowych, w tym raka płuc, piersi i jajników[42,43]. Inhibicja CrkII prowadzi z mechanicznego punktu widzenia do ograniczenia rozbudowy sieci aktynowych mikrofilamentów tworzących szkielet komórkowy, czyli do efektów podobnych do morfologicznych skutków nadekspresji PEAK3 w komórce,

---

[41] Kung JE, Jura N. Prospects for pharmacological targeting of pseudokinases. Nat Rev Drug Discov. 2019 Jul;18(7):501-526. doi: 10.1038/s41573-019-0018-3,PMID: 30850748.

[42] Rodrigues S. P. et al., CrkI and CrkII function as key signaling integrators for migration and invasion of cancer cells. Mol. Cancer Res. 2005:3, 183–194, PMID: 15831672

[43] Nakashima N, Rose DW, Xiao S, Egawa K, Martin SS, Haruta T, Saltiel AR, Olefsky JM. The functional role of CrkII in actin cytoskeleton organization and mitogenesis. J Biol Chem. 1999 Jan 29;274(5):3001-8, PMID: 9915838.

obserwowanych przez Lopeza i in[44]. PEAK3 może okazać się potencjalnym celem w farmakologicznej walce z AML i innymi typami nowotworów, które są podatne na inhibitory szlaków ruchliwości i migracji komórek.

[44] Lopez ML, Lo M, Kung JE, Dudkiewicz M, Jang GM, Von Dollen J, Johnson JR, Krogan NJ, Pawłowski K, Jura N. Proc Natl Acad Sci U S A. 2019 Jul 30;116(31):15495-15504. doi: 10.1073/pnas.1906360116,PMID: 31311869.

II. Podsumowanie aktywności naukowej
[INFORMATION ON SCIENTIFIC OR ARTISTIC ACTIVITY]

1). Lista rozdziałów w monografiach naukowych i podręcznikach.
[List of published chapters in scientific monographs]

Rozdziały w monografiach	
A) międzynarodowych	
Lp.	Opis bibliograficzny pracy*
1	Lange A, Polak M, **Dudkiewicz M,** Kościńska K, Nowicka A, Bogunia-Kubik K. Standardization of Donor-Recipient Matching in Transplantation. Hauppage NY(USA): Nova Science Publishers, Inc., **2006**; ISBN:1-59454-423-9. Section II: NATIONAL BONE MARROW DONOR REGISTRIES; Chapter 11 ACTIVITY OF THE NATIONAL POLISH BONE MARROW DONOR REGISTRY, 93:101;
2	Śliwka P, **Dudkiewicz M.** BIOMAT 2008. Singapore: World Scientific Publishing Co. Pte. Ltd., **2009**, ISBN:978-981-4271-81-3. The Markov Chains (Markov Set-Chains) as a Tool for Bacterial Genomes Evolution Analysis, 235:252
3	Śliwka P, **Dudkiewicz M.** Symmetries and Groups in Contemporary Physics. Singapore: World Scientific Publishing Co. Pte. Ltd., **2013**, ISBN:978-981-4518-54-3. How to identify laterally transferred genes?, 237:242.
4	Robinett P, Anant J, Beksac M, Danielewicz R, Dausend J, **Dudkiewicz M**, Hwang W, Maiers M, Nazaretyan M, Rasche A, Shiriki N, Stein J, Rajagopal R. A gift of life. WMDA handbook for Blood stem cel donation. Leiden: World Marrow Donor Association. **2013**; ISBN:978-90-821221-0-7. Chapter 1. General Organization of a registry, 23:43.
B) krajowych	
Lp.	Opis bibliograficzny pracy*
1	Boguradzki P, Firley H, **Dudkiewicz M**. Pielęgniarstwo Transplantacyjne. Poznań: Wydawnictwo Ars Nova. **2014**; ISBN:978-83-7637-227-3. Rozdział 15. Przeszczepianie szpiku i komórek krwiotwórczych, 307:334.

2). Lista publikacji w czasopismach naukowych, w tym nieuwzględnionych w sekcji I.2., z podziałem na prace opublikowane przed i po uzyskaniu stopnia doktora.
[List of articles published in scientific journals (including the articles not mentioned in section I.2)].

Przed uzyskaniem stopnia doktora

	I. Oryginalne pełnotekstowe prace naukowe (bez streszczeń, zjazdowych i konferencyjnych, prac w suplementach czasopism, listów do redakcji oraz udziału autora wymienionego w dodatku (appendix) jako uczestnika badań wieloośrodkowych)				
	A) w czasopismach posiadających Impact Factor (IF) – chronologicznie				
L.p.	Opis bibliograficzny pracy*	**IF**	**KBN/ MNiSW**	**IC**	**Punktacja MNiSW 2019**
1.	Mackiewicz P, Szczepanik D, Gierlik A, Kowalczuk M, Nowicka A, **Dudkiewicz M**, Dudek MR, Cebrat S. The differential killing of genes by inversions in prokaryotic genomes. J. Mol. Evol. **2001**;53(6):615-621.	4,011	15		
2.	Kowalczuk M, Mackiewicz P, Mackiewicz D, Nowicka A, Dudkiewicz M, Dudek MR, Cebrat S. Multiple base substitution corrections in DNA sequence evolution.. Int. J. Modern Phys. C. 2001;12(7):1043-1053.	0,728	8		

3.	Mackiewicz P, Kowalczuk M, Mackiewicz D, Nowicka A, **Dudkiewicz M**, Łaszkiewicz A, Dudek MR, Cebrat S. How many protein-coding genes are there in the Saccharomyces cerevisiae genome? Yeast. **2002**;19(7):619-629.	2,340	11		
4.	Mackiewicz P, Kowalczuk M, Mackiewicz D, Nowicka A, **Dudkiewicz M**, Łaszkiewicz A, Dudek MR, Cebrat S. Replication associated mutational pressure generating long-range correlation in DNA. Physica A. Statistical Mechanics and its Applications. **2002;**314(1-4):646-654.	1,369	9		
5.	Mackiewicz P, Mackiewicz D, Kowalczuk M, **Dudkiewicz M**, Dudek MR, Cebrat S. High divergence rate of sequences located on different DNA strands in closely related bacterial genomes. Journal of Applied Genetics. **2003**;44(4):561-584	1,725	5		
6.	Mackiewicz P, **Dudkiewicz M**, Kowalczuk M, Mackiewicz D, Banaszak J, Polak N, Smolarczyk K, Nowicka A, Dudek MR, Cebrat S, Differential gene survival under asymmetric directional mutational pressure. Lecture Notes in Computer Science. **2004**;3039: 687-693.	0,513	8		
7.	Polak N, Banaszak J, Mackiewicz P, **Dudkiewicz M**, Kowalczuk M, Mackiewicz D, Smolarczyk K, Nowicka A, Dudek MR, Cebrat S. How Gene Survival Depends on Their Length. Lecture Notes in Computer Science. **2004**;3039: 694-699.	0,513	8		
8.	**Dudkiewicz M**, Mackiewicz P, , Kowalczuk M, Mackiewicz D, Nowicka A, Polak N, Smolarczyk K, Dudek M. R, Cebrat S. Simulation of gene evolution under directional mutational pressure. Physica A. Statistical Mechanics and its Applications. **2004**;336(1-2):63-73.	1,369	9		
9.	**Dudkiewicz M**, Mackiewicz P, Nowicka A, Kowalczuk M, Mackiewicz D, Polak N, Smolarczyk K, Banaszak J, Dudek M. R, Cebrat S. Correspondance between mutation and selection pressure and the genetic code degeneracy in the gene evolution. Future Generation CS. **2005**;21(7):1033-1039.	0,555	15		
	Liczba prac: 9				
Liczba punktów:		13,12	88		

	B) w czasopismach nie posiadających Impact Factor (IF) – chronologicznie			
L.p.	Opis bibliograficzny pracy*	**KBN/ MNiSW**	**IC**	
1.	Kowalczuk M, Mackiewicz P, Mackiewicz D, Nowicka A, Dudkiewicz M, Dudek MR, Cebrat S. DNA asymmetry and the replicational mutational pressure. Journal of Applied Genetics. 2001;42(4):553-577.	5	5,11	
2.	Kowalczuk M, Mackiewicz P, Mackiewicz D, Nowicka A, **Dudkiewicz M**, Dudek MR, Cebrat S. High correlation between the turnover of nucleotides under mutational pressure and DNA composition. BMC Evolutionary Biology. 2001;1:13-18.	6		
3.	Mackiewicz D, Mackiewicz P, Kowalczuk M, **Dudkiewicz M**, Dudek MR, Cebrat S. Rearrangements between Differently Replicating DNA strands in Asymmetric Bacterial Genomes. Acta Microbiologica Polonica. 2003;52(3):245-261.	5	7,01	
4.	**Dudkiewicz M**, Mackiewicz P, Nowicka A, Kowalczuk M, Mackiewicz D, Polak N, Smolarczyk K, Dudek MR, Cebrat S. Properities of the genetic code under directional, asymmetric mutational pressure. Lecture Notes in Computer Science. 2003;2657:343-350	-		-
	Liczba prac: 4			
Liczba punktów:		16	12,1	

Po uzyskaniu stopnia doktora:

I. Oryginalne pełnotekstowe prace naukowe

(bez streszczeń, zjazdowych i konferencyjnych, prac w suplementach czasopism, listów do redakcji oraz udziału autora wymienionego w dodatku (appendix) jako uczestnika badań wieloośrodkowych)

A) w czasopismach posiadających Impact Factor (IF) – chronologicznie

L.p.	Opis bibliograficzny pracy*	IF	KBN/ MNiSW	IC	Punktacja MNiSW 2019
1.	**Dudkiewicz M**, Mackiewicz P, Kowalczuk M, Mackiewicz D, Nowicka A, Polak N, Smolarczyk K, Dudek MR, Cebrat S. Higher mutation rate helps to rescue genes from the elimination by selection. Biosystems. **2005**;80(2):193-199	1,144	15		
2.	**Dudkiewicz M**, Siminska J, Pawlowski K, Orzechowski S. Bioinformatics Analysis of Oligosaccharide Phosphorylation Effect on the Stabilization of the β-Amylase Ligand Complex. Journal of Carbohydrate Chemistry. **2008**;27(8 & 9):479 – 495.	1,144	20		
3.	Jakubska-Busse A, **Dudkiewicz M**, Jankowski P, Sikora R. Mathematical inference of the underground clonal growth of *Epipactis helleborine* (L.) Crantz (*Orchidaceae, Neottieae)*. Bot. Helv. **2009**;119:69–76.	0,900	20		
4.	Czerwiński J, Antoszkiewicz K, Pszenny A, Malanowski P, **Dudkiewicz M**, Woderska A, Kobus G, Ciszek M, Wałaszewski J. Present Data on Organ Donation and Transplantation in Poland. Transplantation Proceedings. **2009**;41:2955–2958.	0,994	20		
5.	**Dudkiewicz M**, Malanowski P, Czerwiński J, Pawłowski K. An Approach to Predicting Hematopoietic Stem Cell Transplantation Outcome Using HLA-Mismatch Information Mapped on Protein Structure Data. Biology of Blood and Marrow Transplantation. 2009, (15), pp.1014-1025.	3,149	30		
6.	Piszczek E, **Dudkiewicz M**, Sobczak M. Molecular cloning and structure modelling analysis of cereal type II metacaspase cDNA from wheat (*Triticum aestivum)*. Biologia Plantarum. **2011**;55(4):614-624.	1,974	25		
7.	Piszczek E, **Dudkiewicz M**, Mielecki M. Biochemical and Bioinformatic Characterization of Type II Metacaspase Protein (TaeMCAII) from Wheat. Plant Mol Biol Rep. **2012**;30(6):1338–1347.	5,319	20		
8.	**Dudkiewicz M**. and Piszczek E. Bacterial putative metacaspase structure from *Geobacter sulfureducens* as a template for homology modeling of type II *Triticum aestivum* metacaspase (TaeMCAII). Acta Biochimica Polonica. **2012**;59(3):1338–1347.	1,185	15		
9.	**Dudkiewicz M**, Szczepińska T, Grynberg M, Pawłowski K.A Novel Protein Kinase-Like Domain in a Selenoprotein, Widespread in the Tree of Life. PLoS ONE. **2012**;7(2): e32138.	3,730	40		
10.	Orzechowski S, Grabowska A, Sitnicka D, Siminska J, Felus M, **Dudkiewicz M**, Fudali S, Sobczak M. Analysis of the expression, subcellular and tissue localisation of phosphoglucan, water dikinase (PWD/GWD3) in *Solanum tuberosum* L.: a bioinformatics approach for the comparative analysis of two a-glucan, water dikinases (GWDs) from *Solanum tuberosum* L. Acta Physiol Plant. **2013**;35:483–500.	1,524	25		

11.	Dudkiewicz M, Lenart A, Pawłowski K. A Novel Predicted Calcium-Regulated Kinase Family Implicated in Neurological Disorders. PLOS ONE. 2013, 8(6), pp. e66427.	3,534	40		
12.	Lenart A, **Dudkiewicz M**, Grynberg M, Pawłowski K. CLCAs - A Family of Metalloproteases of Intriguing Phylogenetic Distribution and with Cases of Substituted Catalytic Sites. PLoS ONE. **2013**;8(5): e62272.	3,534	40		
13.	Kodavali PK, Dudkiewicz M, Pikuła S, Pawłowski K. Bioinformatics Analysis of Bacterial Annexins – Putative Ancestral Relatives of Eukaryotic Annexins. PLoS ONE. **2014**, 9(1), e85428.	3,234	40		
14.	Nguyen KB, Sreelatha A, Durrant ES, Lopez-Garrido J, Muszewska A, **Dudkiewicz M**, Grynberg M, Yee S, Pogliano K, Tomchick DR, Pawłowski K, Dixon JE, Tagliabracci V. Phosphorylation of spore coat proteins by a family of atypical protein kinases. Proc Natl Acad Sci U S A. **2016**;113(25):E3482-91.	9.661	45		
15.	Czerwiński J, Antoszkiewicz K, Grygiel K, Karpeta E, Górski Ł, Dudkiewicz M, Lewandowska D. National Transplants Registry in Poland: early and Long-term Results of Organ Transplantations in the Years 1998 to 2014. Transpl. Proc. **2016**;48: 1407-1410.	0,908	15		
16.	Łęczycka A, Dudkiewicz M, Czerwiński J, Malanowski P, Żalikowska-Hołoweńko J, Danielewicz R. National Hematopoietic Stem Cells Transplant Registry in Poland: Nationwide Internet Reporting System and Results. Transpl. Proc. **2016**;48: 1791-1796.	0,908	15		
17.	Hareza A, Bakun M, Świderska B, Dudkiewicz M, Koscielny A, Bajur A, Jaworski J, Dadlez M, Pawłowski K. Phosphoproteomic insights into processes influenced by the kinase-like protein DIA1/C3orf58. PeerJ. **2018** Apr 9(6):e4599. doi:10.7717/peerj.4599.	2,35	35		
18.	Lopez ML, Lo M, Kung JE, Dudkiewicz M, Jang GM, Von Dollen J, Johnson JR, Krogan NJ, Pawłowski K, Jura N. PEAK3/C19orf35 pseudokinase, a new NFK3 kinase family member, inhibits CrkII through dimerization Proc Natl Acad Sci U S A. **2019**;116(31): 15495-15504.	9,580			200
19.	Dudkiewicz M, Pawłowski K. A novel conserved family of Macro-like domains-putative new players in ADP -ribosylation signaling. PeerJ. **2019**;1(7):e6863.	2,353			100
20	Nestorowicz K, Bogacz A, Bukowska A, Chraplak M, Czerwiński J, Góralski M, Gronkowski M, Jopek K, Kniżewski Ł, Kolasiński M, Kowalski ML, Nowak J, Sowiński M, Wróblewska-Kabba S, Tymoniuk B, Dudkiewicz M. High-resolution allele frequencies for NGS based HLA-A, B, C, DQB1 and DRB1 typing of 23 595 bone marrow donors recruited for the Polish central potential unrelated bone marrow donor registry. Hum Immunol. **2020**;81(2-3):49-51.	2,202			70
21	Hurley CK, Kempenich J, Wadsworth K, Sauter J, Hofmann JA, Schefzyk D, Schmidt AH, Galarza P, Cardozo MBR, Dudkiewicz M, Houdova L, Jindra P, Sorensen BS, Jagannathan L, Mathur A, Linjama T, Torosian T, Freudenberger R, Manolis A, Mavrommatis J, Cereb N, Manor S, Shriki N, Sacchi N, Ameen R, Fisher R, Dunckley H, Andersen I, Alaskar A, Alzahrani M, Hajeer A, Jawdat D, Nicoloso G, Kupatawintu P, Cho L, Kaur A, Bengtsson M, Dehn J. Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0.0. HLA. **2020** Jan 22, 95(6): 516-531.	2,785			70

22.	Nestorowicz K, **Dudkiewicz M**, Czerwiński J, Łęczycka A, Macher M, Kamiński A. Central Unrelated Potential Bone Marrow and Cord Blood Registry in Poland: Structure and Numbers. Transplant Proc. **2020**;S0041-1345(19)31795-6.	0,959			40
Liczba prac: 22					
Liczba punktów:		**63,07**	**460**		**480**
B) w czasopismach nieposiadających Impact Factor (IF) – chronologicznie					
L.p.	Opis bibliograficzny pracy*		**KBN/ MNiSW**	**IC**	
1.	Dudek MR, Cebrat S, Kowalczuk M, Mackiewicz P, Nowicka A, Mackiewicz D, **Dudkiewicz M.** Information Weights of Nucleotides in DNA Sequences. Computational Methods in Science and Technology. **2007**; 13(1):5-12.		9		-
2.	Jakubska A, **Dudkiewicz M**, Jankowski P, Sikora R. A mathematical method to estimate the underground development of the clonal species. Colloquium Biometricum. **2008**;38:185-192.		6		-
Liczba prac: 2					
Liczba punktów:			**15**		

3). Udział i prezentacje na konferencjach krajowych i międzynarodowych.
[Information on presentations given at national or international scientific or arts conferences, including a list of lectures delivered upon invitation and plenary lectures].

2008
BIOMAT 2008, International Symposium on Mathematical and Computational Biology, Campos do Jordão, Brazil, 22 – 27 November 2008, prezentacja ustna:
- The Markov Chains (Markov Set-Chains) as a Tool for Bacterial Genomes Evolution Analysis.

33rd FEBS Congress and 11th IUBMB Conference, June 28-July 3, 2008, Athens, Greece, prezentacje posterowe:
- USING BIOINFORMATICS FOR ANALYSIS OF HLA MISMATCHES IN DONOR-RECIPIENT MATCHING PROCESS AT HSCT TRANSPLANTATION.
- Bioinformatics analysis of oligosaccharides phosphorylation influence on the stabilization of the beta-amylase-ligand complex.

2009
ISMB/ECCB 2009- 17th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) and 8th Conference on Computational Biology (ECCB)- Stockholm, June 27-July 2 2009 : prezentacje posterowe:
- AN APPROACH TO PREDICTION OF THE HAEMATOPOIETIC STEM CELL TRANSPLANTATION OUTCOME USING HUMAN LEUKOCYTE ANTIGEN MISMATCH
- INFORMATION MAPPED ON PROTEIN STRUCTURE DATA ON EXAMPLES OF HLA-B AND HLA-C ANTIGENS

2010
Experimental approaches to protein-protein interactions, 11-12 January 2010, Sheffield, UK, prezentacja posterowa:
- PROTEIN-PROTEIN INTERACTIONS BETWEEN STARCH RELATED ENZYMES IN POTATO

The EMBO Meeting 2010-Barcelona , September 4 – 7 2010, : prezentacje posterowe:
- BIOINFORMATICS ANALYSIS OF HLA-Cp/KIR COMPLEXES AND THEIR INFLUENCE ON HEMATOPOIETIC STEM CELL TRANSPLANTATION OUTCOME.
- EVOLUTION OF APPARENTLY INACTIVE ACTIVE SITES IN THE HEXXH METALLOPROTEASES.

2011
The 2nd Polish Congress of Biochemistry and Cell Biology, 5-9 September 2011, Kraków, prezentacja posterowa:
- Homology modeling of type II Triticum aestivum metacaspase (TaeMCAII) using a bacterial putative metacaspase structure from Geobacter sulfureducens.

EPGRC 2011, EUCARPIA Meeting, 5-7 April 2011, Wageningen, The Netherlands, prezentacja posterowa:
- Bioinformatics approach to the strawberry transcriptome analysis in the context of the response to fungal pathogen.

2012
Group Theoretical Methods in Theoretical Physics, ICGTMP "Gr-29". (20-26) August 2012 Tianjin, sesja Application in Biology: prezentacja ustna:

- How to identify laterally transferred genes? Mathematical and statistical approach to HGT phenomenon between Prokaryota and Eukaryota.

2013
39th Meeting of European Group of Blood and Marrow Transplantation, 7-10 April 2013 London, UK, prezentacja posterowa:

- First outcome of combined public/private umbilical cord blood banking in Poland.

Bioinformatics 2013/BIT 2013, 26-29 June 2013, Toruń, prezentacja posterowa:

- Identification of novel HExxH metaloprotease domains in the proteome grey zones using bioinformatics methods

2014
XI Sympozjum Polskiego Towarzystwa Transplantacyjnego w Poznaniu, 26-27 września 2014, prezentacja ustna:

- Centralny Rejestr Niespokrewnionych Potencjalnych Dawców Szpiku i Krwi Pępowinowej + organizacja systemu przeszczepiania komórek krwiotwórczych.

Prezentacje posterowe:

- Rejestr przeszczepień komórek krwiotwórczych szpiku i krwi obwodowej oraz krwi pępowinowej.
- Centralny Rejestr Niespokrewnionych Potencjalnych Dawców Szpiku i Krwi Pępowinowej- organizacja systemu przeszczepiania komórek krwiotwórczych w Polsce.

WMDA Fall Meeting/NMDP Council Meeting, 5-8 November 2014, Minneapolis USA, prezentacja posterowa:

- Effect of increasing of national donors' pool in Poland on Polish patients and their chance for finding unrelated donor.

2015
WMDA Search Coordinator Certificate Training Programme Basic Level (October 2015- June 2016).
2016
8 Kurs Transplantologii Praktycznej, Klinika Immunologii i Chorób Wewnętrznych Instytutu Transplantologii Warszawskiego Uniwersytetu Medycznego, 22 -23 stycznia 2016, Warszawa, prezentacja ustna:

- Rejestr przeszczepień: wyniki przeszczepiania komórek krwiotwórczych w Polsce.

WMDA Search Coordinator Certificate Training Programme Advanced Level (October 2016- June 2017)
World Marrow Donor Association spring meeting and International Donor Registries Conference, Singapore, 30 May-2 June 2016, prezentacje posterowe:

- 1 million donors registry in Poland?
- Does increasing national donors' pool effect chance for finding matched unrelated donor for Polish patients?

VI Międzynarodowa Konferencja Naukowo-Szkoleniowa Transplantacja Komórek Krwiotwórczych od dawców haploidentycznych, 14-15 października 2016, Poznań, prezentacja ustna:

- Omówienie bieżących zagadnień związanych z koordynacja transplantacji allogenicznych komórek krwiotwórczych w Polsce.

2017
Sympozjum z okazji 150 lecia Lekarzy Polskich we Lwowie, Lwów 18-21 października 2017 r., prezentacja ustna:

- Organizacja i rola Rejestru Niespokrewnionych Potencjalnych Dawców Szpiku i Krwi Pępowinowej

2018
VIII Międzynarodowa Konferencja Naukowo-Szkoleniowa Transplantacja Komórek Krwiotwórczych w leczeniu zespołów mielodysplastycznych i ostrych białaczek szpikowych, 12-13 października 2018, Poznań, prezentacja ustna:

- GRID-Global Registry Identifier for donor.

2019
IPK2019, Inhibitors for Protein Kinases Conference, Warszawa 14-18 września 2019, prezentacje posterowe

4). Udział w komitetach organizacyjnych i naukowych krajowych lub międzynarodowych konferencji

[Information on participation in organizational and scientific committees at national or international conferences, including the applicant's function]

Szkolenie w ramach Narodowego programu Rozwoju Medycyny Transplantacyjnej: „Organizacja współpracy i leczenie z zastosowaniem hematopoetycznych komórek macierzystych", 10-12 grudnia 2015 Błonie, prezentacja ustna:

- Ustawowa działalność Centralnego Rejestru Niespokrewnionych Potencjalnych Dawców Szpiku i Krwi Pępowinowej oraz ustawowych Rejestrów Transplantacyjnych.

I Konferencja Szkoleniowa Koordynatorów pobierania i przeszczepiania komórek krwiotwórczych w ośrodkach hematologicznych, Warszawa, 27-28 listopad 2017 r. –

- Organizacja i prowadzenie wykładów/warsztatów

Szkolenie „Organizacja współpracy i leczenie z zastosowaniem hematopoetycznych komórek macierzystych" dla personelu ośrodków pobierających komórki krwiotwórcze, banków komórek krwiotwórczych, ośrodków stosujących komórki krwiotwórcze, ośrodków testującychi ośrodków dawców szpiku", Zegrze, 4-5 grudnia 2017

- Z perspektywy Poltransplantu: Współpraca z ośrodkami dawców szpiku, dobierającymi, przeszczepiającymi i rejestrami

II Konferencja Szkoleniowa Koordynatorów pobierania i przeszczepiania komórek krwiotwórczych w ośrodkach hematologicznych, Warszawa, 12-13 listopad 2018 r.

- Organizacja i prowadzenie wykładów/warsztatów

III Konferencja szkoleniowa koordynatorów pobierania i przeszczepiania komórek krwiotwórczych w ośrodkach hematologicznych, Warszawa, 7-8 listopada 2019 r

- Organizacja i prowadzenie wykładów/warsztatów

5). Udział w realizacji grantów naukowo-badawczych

[Information on participation in the works of research teams realizing projects financed through national and international competitions, including the projects which have been completed and projects in progress, and information on the function performed in the team.]

2008-2009 - **kierownik projektu** badawczego własnego „Modelowanie alloreaktywności ludzkich receptorów limfocytów T - badanie kompleksów TCR/pMHC w kontekście zmienności antygenów HLA pod kątem zastosowania w transplantologii". (Katedra Doświadczalnictwa i Bioinformatyki WRiB SGGW)

2007-2010 - **wykonawca** projektu badawczego własnego „Analiza *in silico* struktur, funkcji i ewolucji białek CLCA - nowej rodziny ludzkich białek o potencjalnych zastosowaniach terapeutycznych". Kierownik projektu: dr hab. Krzysztof Pawłowski (Instytut Biologii Eksperymentalnej PAN im. M. Nenckiego).

2009-2011 **wykonawca** projektu badawczego własnego „Analiza biochemiczna i strukturalna metakaspazy typu II z pszenicy (*Triticum aestivum*)". Kierownik projektu: dr Ewa Piszczek (Katedra Biochemii WRiB SGGW).

2015-2018 **wykonawca** projektu badawczego własnego „Znaczenie izolacji geograficznej Ameryki Południowej w ewolucji bakterii brodawkowych z rodzaju *Bradyrhizobium* na obszarze Krainy Neotropikalnej". Kierownik projektu: dr hab. Tomasz Michał Stępkowski (Samodzielny Zakład Biologii Mikroorganizmów, WRiB SGGW).

III. Wdrożenia:

[INFORMATION ON COOPERATION WITH SOCIAL AND ECONOMIC ENVIRONMENT]

2010-2019 udział w projektowaniu, tworzeniu i rozwoju działającej obecnie wersji systemu informatycznego dla Centralnego Rejestru Potencjalnych Niespokrewnionych Dawców Szpiku i Krwi Pępowinowej, administrowanie bazą danych, nadzorowanie aktualizacji i walidacji, czuwanie nad integralnością danych i funkcjonowaniem procesów i połączeń międzysystemowych.

2018- 2020: współpraca z Centrum Systemów Informatycznych Ochrony Zdrowia w zakresie projektowania i wdrożenia nowego systemu informatycznego dla polskiej transplantologii, koordynacja pracy zespołu ds. algorytmu wyszukiwania i doboru potencjalnie zgodnych dawców komórek krwiotwórczych z uwzględnieniem częstości haplotypów HLA w populacji polskiej.

IV. Podsumowanie łącznego dorobku naukowego
[SCIENTOMETRIC INFORMATION]

	Przed doktoratem		Po doktoracie		
	IF	**MNiSW/KBN**	**IF**	**MNiSW/KBN**	**Punktacja MNiSW 2019***
Oryginalne pełnotekstowe prace naukowe	13,12	104	63,071	475	480

Łącznie:

- **IF = 75,19**
- **MNiSW/KBN= 579+480***
- Liczba cytowań z bazy Scopus z dn. 30.07.2020 r. = **362**
- Index Hirscha z bazy Scopus z dn. 30.07.2020 r. = **11**

*wg Wykazu czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych wraz z przypisaną liczbą punktów z dnia 31 lipca 2019 r.

5. Aktywność naukowa realizowana w więcej niż jednej uczelni, instytucji naukowej, w szczególności zagranicznej oraz staże naukowe

[Presentation of significant scientific or artistic activity carried out at more than one university, scientific or cultural institution, especially at foreign institutions]

Publikacje tworzące dorobek naukowy powstawały w skutek aktywności naukowej realizowanej w następujących ośrodkach i instytucjach:

Okres przed uzyskaniem stopnia doktora:

Uniwersytet Wrocławski, Wydział Nauk Przyrodniczych
Po uzyskaniu stopnia doktora:

-Dolnośląskie Centrum Transplantacji Komórkowych z Krajowym Bankiem Dawców Szpiku
-Wydział Rolnictwa i Biologii SGGW w Warszawie,

- Centrum Koordynacyjno-Organizacyjne ds. Transplantacji POLTRANSPLANT
Praca w Centralnym Rejestrze Niespokrewnionych Potencjalnych Dawców Szpiku oparta jest na stałych kontaktach z instytutami i uniwersytetami medycznymi w kraju (m.in. WUM i UCK w Warszawie, UCK w Gdańsku, Instytut Hematologii i Transfuzjologii w Warszawie) i za granicą – WMDA (World Marrow Donor Association), NMDP (National Marrow Donor Program USA) oraz Center for International Blood and Marrow Transplant Research (CIBMTR), Medical College of Wisconsin. Obowiązki kierownika zespołu Centralnego Rejestru wymagają licznych wyjazdów zagranicznych i międzynarodowych szkoleń oraz udziału w pracach międzynarodowych zespołów badawczych w zakresie populacyjnych badań nad częstościami alleli i haplotypów zgodności tkankowej.

Publikacje powstałe w wyniku prowadzenia badań w więcej niż jednej jednostce naukowej:

-W ramach współpracy z Instytutem Biologii Doświadczalnej im. M. Nenckiego PAN w Warszawie:

Dudkiewicz M, Lenart A, Pawłowski K. A Novel Predicted Calcium-Regulated Kinase Family Implicated in Neurological Disorders. PLoS ONE. 2013;8(6): e66427.

Lenart A, Dudkiewicz M, Grynberg M, Pawłowski K. CLCAs - A Family of Metalloproteases of Intriguing Phylogenetic Distribution and with Cases of Substituted Catalytic Sites. PLoS ONE. 2013;8(5): e62272.

Kodavali PK, Dudkiewicz M, Pikuła S, Pawłowski K. Bioinformatics analysis of bacterial annexins, putative ancestral relatives of eukaryotic annexins. PLoS ONE. 2014;9(1):e85428.

W ramach współpracy z Department of Pharmacology, University of California, San Diego, La Jolla oraz Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas:

Nguyen KB, Sreelatha A, Durrant ES, Lopez-Garrido J, Muszewska A, Dudkiewicz M, Grynberg M, Yee S, Pogliano K, Tomchick DR, Pawłowski K, Dixon JE, Tagliabracci V. Phosphorylation of spore coat proteins by a family of atypical protein kinases. Proc Natl Acad Sci U S A. 2016;113(25):E3482-91; doi:10.1073/pnas.1605917113.

-W ramach współpracy z Instytutem Biochemii i Biofizyki PAN w Warszawie:

Hareza A, Bakun M, Świderska B, Dudkiewicz M, Koscielny A, Bajur A, Jaworski J, Dadlez M, Pawłowski K. Phosphoproteomic insights into processes influenced by the kinase-like protein DIA1/C3orf58. PeerJ. 2018 Apr 9(6):e4599. doi:10.7717/peerj.4599.

-W ramach współpracy z University of California, San Francisco, CA:

Lopez ML, Lo M, Kung JE, Dudkiewicz M, Jang GM, Von Dollen J, Johnson JR, Krogan NJ, Pawłowski K, Jura N. PEAK3/C19orf35 pseudokinase, a new NFK3 kinase family member, inhibits CrkII through dimerization. Proc Natl Acad Sci U S A. 2019;116(31): 15495-15504.

-W ramach współpracy z Uniwersytetem Medycznym w Warszawie oraz Centrum Koordynacyjno-Organizacyjnym ds. Transplantacji Poltransplant

Czerwiński J, Antoszkiewicz K, Grygiel K, Karpeta E, Górski Ł, Dudkiewicz M, Lewandowska D. National Transplants Registry in Poland: early and Long-term Results of Organ Transplantations in the Years 1998 to 2014. Transpl. Proc. 2016;48: 1407-1410.

Łęczycka A, Dudkiewicz M, Czerwiński J, Malanowski P, Żalikowska-Hołoweńko J, Danielewicz R. National Hematopoietic Stem Cells Transplant Registry in Poland: Nationwide Internet Reporting System and Results. Transpl. Proc. 2016;48: 1791-1796.

Nestorowicz K, Bogacz A, Bukowska A, Chraplak M, Czerwiński J, Góralski M, Gronkowski M, Jopek K, Kniżewski Ł, Kolasiński M, Kowalski ML, Nowak J, Sowiński M, Wróblewska-Kabba S, Tymoniuk B, Dudkiewicz M. High-resolution allele frequencies for NGS based HLA-A, B, C, DQB1 and DRB1 typing of 23 595 bone marrow donors recruited for the Polish central potential unrelated bone marrow donor registry. Hum Immunol. 2020 Jan 2. pii: S0198-8859(19)31234-0. doi: 10.1016/j.humimm.2019.12.005. [Epub ahead of print].

Hurley CK, Kempenich J, Wadsworth K, Sauter J, Hofmann JA, Schefzyk D, Schmidt AH, Galarza P, Cardozo MBR, Dudkiewicz M, Houdova L, Jindra P, Sorensen BS, Jagannathan L, Mathur A, Linjama T, Torosian T, Freudenberger R, Manolis A, Mavrommatis J, Cereb N, Manor S, Shriki N, Sacchi N, Ameen R, Fisher R, Dunckley H, Andersen I, Alaskar A, Alzahrani M, Hajeer A, Jawdat D, Nicoloso G, Kupatawintu P, Cho L, Kaur A, Bengtsson M, Dehn J. Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0.0. HLA. 2020. Jan 22. doi: 10.1111/tan.13811. [Epub ahead of print].

6. Osiągnięcia dydaktyczne:

[Presentation of teaching and organizational achievements as well as achievements in popularization of science or art.]

-opracowanie programu nowego przedmiotu obejmującego wykłady i ćwiczenia dla studentów IV roku biologii Wydziału Rolnictwa i Biologii SGGW: „Bioinformatyka", prowadzenie zajęć w latach 2007-2020.

-opracowanie programu i prowadzenie zajęć z przedmiotów: „Analiza genetyczna", „Filogenetyka molekularna", „Genomika i proteomika cyfrowa" w ramach studiów podyplomowych Biostatystyka i analiza danych w ochronie zdrowia publicznego (2015/2016).

-dziesięcioletnie doświadczenie w prowadzeniu zajęć z przedmiotów: Zastosowanie komputerów, Informatyka, Podstawy informatyki, Planowanie eksperymentu i statystyka, Matematyka i statystyka matematyczna, Statystyka, Metody statystyczne w biologii dla studentów Wydziału Rolnictwa i Biologii oraz Technologii Żywności SGGW.

- prace dyplomowe i doktorskie:

Promotor pracy magisterskiej: IB, Katedra Biochemii i Mikrobiologii: „Bioinformatyczna charakterystyka efektorów *Legionella pneumophila* zawierających przypuszczalnie domeny pochodzenia eukariotycznego".

Promotor pomocniczy prac doktorskich:

Katedra Biochemii i Mikrobiologii IB SGGW: „Identyfikacja nowych, nieadnotowanych domen ADP-rybozylotransferaz w bazach danych sekwencyjnych".

Warszawski Uniwersytet Medyczny: „Immunogenetyczne, kliniczne i organizacyjne wyznaczniki doboru niespokrewnionych dawców przeszczepów komórek krwiotwórczych dla chorych z chorobami hematologicznymi."

7. Inne informacje

[Apart from information set out in 1-6 above, the applicant may include other information about his/her professional career, which he/she deems important.]

Nagrody i odznaczenia:
2016 - Odznaka honorowa Ministerstwa Zdrowia "Za zasługi dla ochrony zdrowia".
2015 - *Pro Transplantationibus Fovendis* – nagroda Polskiego Towarzystwa Transplantacyjnego.
2010 - Nagroda Rektora SGGW za osiągnięcia naukowe, indywidualna III stopnia

30.07.2020
.....
data

.....
Podpis Wnioskodawcy